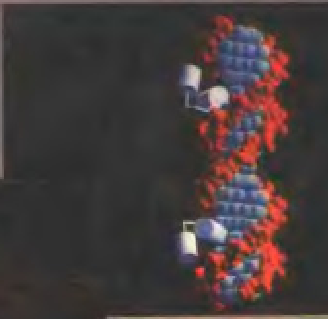
高等学校试用教材

# Molecular Genetic

# 分子遗传学

· 孙乃恩
· 孙东旭 编著
· 朱德煦



南京大学出版社

高等学校试用教材

# 分子遗传学

孙乃恩　孙东旭　朱德煦　编著

南京大学出版社

1995·南京

(苏)新登字011号

**内 容 提 要**

本书内容丰富、新颖、文字流畅，可读性强，包括DNA的结构，基因组和染色体，DNA复制，修复，突变，转录，翻译，原核调控，真核调控，调控的分子机制，遗传重组，基因工程等12章。每章后附有参考文献(全书共700余篇)。

该书已由国家教委遗传学教材编审组正式审定通过，并推荐为高等院校有关专业作为试用教材。可供综合性大学、师范、农林、医药院校有关专业本科生、研究生及教师使用，对于从事生命科学的研究人员也是一本有益的参考书。

**分 子 遗 传 学**

孙乃恩 孙东旭 朱德煦 编著

*

南京大学出版社出版

(南京大学校内)

江苏省新华书店发行　　丹阳新华印刷厂印刷

*

开本：787×1092 1/16 印张：33.25 字数：826千

1990年8月第1版　　1995年1月第4次印刷

印数：5501—9500

ISBN 7-305-00692-0/Q·5

定价：21.00元

责任编辑 荣兴琴

# 序

分子遗传学是在分子水平上研究生物遗传和变异的遗传学分支学科。它经历了近40年的突飞猛进的发展，已成为生命科学的前沿学科之一；它的基本理论已渗透到生命科学的几乎所有领域，促进了一大批新学科和新技术的兴起和发展，令人瞩目。

我国从八十年代开始，许多高等学校相继开设了分子遗传学课程，并积极进行教材建设，编写了一批各有侧重和特色的教材。1988年6月在杭州召开的全国理科遗传学教材评审会议上，大家认为孙乃恩等同志编著的《分子遗传学》内容翔实而系统，反映了当前国际上分子遗传学的新进展，比较适合于我国目前多数学校的教学需要。

该书作者严谨认真，结合自己的教学和科研，阅读了大量的文献资料。他们不断地追踪国际上最新进展，为该书增添新的血液，书中竟引用了1990年的国外新资料。该书注重分子遗传学的基本内容，阐发其基本原理，并以此为中心将收集的资料去粗取精。作者强调生物大分子结构和功能的关系，刻意安排了"遗传的物质基础——DNA"作为第一章。其内容和分量恰到好处，既弥补了一般生物学教科书的不足之处，又不令人望而生畏。全书内容大体上按照遗传信息的传递过程进行安排，顺其自然。最后，以天然的遗传重组和人工遗传重组（基因工程）结束全书，让读者认识到，遗传信息并不是僵硬地和一成不变地排列在各个独立的染色体上，而是处于某种程度的流动状态。这样既扩展了读者对生命起源和进化过程的认识，又概括地向读者展现了人类改造生命与创造新生命的伟大业绩。

该书内容新颖，语言流畅。因此，我很乐意向读者推荐这本好书，希望它能有助于我国分子遗传学教学水平的进一步提高，并对生命科学研究工作者有所裨益。

陈 士 怡

1990 年 7 月 2 日于杭州

# 前　　言

在我们居住的地球上，有大约 1000 万种生物。有的生物只是一个单细胞，如大肠杆菌和酵母菌；有的则有复杂的器官和组织，包含着大量的具有不同的结构、形态和功能的细胞，像人体就有 $10^{14}$ 个细胞。如果就生物大分子而言，人体有大约 50 000 种以上的蛋白质，同时含有数以万计的核酸及其它大分子种类。地球上的全部生物，估计包含 $10^{11}$ 种蛋白质和差不多相同数量的核酸种类。即使是极为简单的大肠杆菌(其体积约为 $2\times10^{-12}cm^3$)，也含有 3 000 多种蛋白质，1 000 多种核酸，还有 1 000 多种其他生物大分子和低分子量的有机化合物。

在这样种类复杂、形态万千的生物体系中，人们必须寻求生命状态的基本逻辑原理，这就是：

(1) 生物大分子虽然具有复杂的结构，但在其组成方面却存在着一种基本的简单性，例如 DNA 由四种脱氧核糖核苷酸聚合而成，RNA 由四种核糖核苷酸聚合而成，蛋白质由 20 余种氨基酸聚合而成，多糖由少数几种单糖聚合而成。

(2) 所有的生物都使用相同种类的构件分子，似乎它们是从一个共同的祖先进化而来。

(3) 每个物种的特性是通过它具有的一套与众不同的核酸和蛋白质而保持的。

(4) 每种生物大分子在细胞中有特定的功能。

以上这些正是生物化学、分子生物学、分子遗传学所要研究的基本问题所遵循的逻辑。

分子遗传学是一门新兴的科学，它可以追溯到 1944 年 Avery 等人的肺炎球菌转化实验；但是突破性的事件乃是 1953 年 Watson 和 Crick 的 DNA 双螺旋结构模型的提出，从而开创了分子生物学和分子遗传学的新纪元。1977 年，Sanger 弄清了噬菌体 $\phi\times174$ 的全部碱基序列(5 386 个碱基)，确立了 DNA 序列分析的新战略和新方法，从而使分子生物学和分子遗传学进入了一个崭新的时代。30 多年来，分子遗传学取得了极其巨大的成就，它已成为生命科学的带头科学之一，有力地促进了生命科学中各分枝学科的发展。在实践方面，分子遗传学促进了基因工程的发展，使得人类能够以更快的速度和更准确的目标进行生物品种的改良，甚至创造新物种，大规模进行生物活性物质的生产。在工业、农业、医药、环境保护等方面，基因工程都已崭露头角，应用范围日益扩大，它正在成为一个有巨大潜力的产业部门。

然而，分子遗传学仍然是一门年轻的学科。一方面，它的发展极为迅速，任何一本教科书都不可能将最新的研究成果毫无遗漏地搜集起来。同时，也正因为其发展迅速，使得人们不得不重新考虑某些原来被公认为是正确的结论。人们对所谓“内含子”和“外显子”的认识就是一个很好的例子。70 年代末刚发现 intron 时，人们普遍认为它是真核基因的“真质标记”(hallmark)，是一种不表达成蛋白质的(即内含而不显的)核酸序列，因而中译名把它叫做“内含子”。可是时过不久，人们就发现酵母线粒体细胞色素 b 基因的一个 intron 是编码一种叫做成熟酶(maturase)的蛋白质的，而且这个蛋白质是细胞色素 b mRNA 前体的拼接过程所不可缺少的。关于 exon，通常真核生物基因的首尾两个 exon 也只有一部分序列编码蛋白质；后来又相继发现若干基因的首尾 exon 完全不编码蛋白质的情况(例如人类尿激酶原基因的第一个 exon 的 88 个核苷酸全部不编码蛋白质)。这样，中译名“内含子”和“外显子”都暴露

出其画蛇添足之嫌。1984 年,Chu 等人发现了 T4 噬菌体胸腺嘧啶核苷酸合成酶的基因中有一个 1017 个核苷酸的 intron，这就使得 intron 是真核基因的真质标记之说从此告终。另一方面,人们对许多遗传现象的分子机制仍然不清楚甚至毫无所知；有些正在探索之中,常会有歧义的甚至完全矛盾的说法。在这种情况下,作者尽量采用流行的说法。

本书原稿是作者在南京大学任课所使用的讲义。它经过七年试用并进行三次大的修改，方成该书。在成书过程中，始终得到各方面的鼓励与支持。特别是全国遗传学教材编审组和高等教育出版社组织的杭州遗传学教材会议(1988 年 7 月)通过对该书的评审。会上，以陈士怡教授为首的评委们提出了宝贵的评审意见并很快转达给作者。作者根据这一意见又作了一次修改。1989 年 11 月的杭州遗传学教材会议又督促该书及早问世。高等教育出版社李茂国同志,杭州大学李桃生同志,南京大学出版社责任编辑荣翠琴同志都为本书的问世付出了辛勤的劳动。对于大家的热情帮助和支持,作者均致以衷心的感谢。

本书以 David Freifelder 的 "Molecular Biology" (1983 年版)，Benjamin Lewin 的 "Genes" (1985 年第二版) 和 James Watson 等人的 "Molecular Biology Of The Gene" (1987 年第四版)为主要参考书,同时吸取一些国外近年 (1986 年—1990 年初) 期刊杂志上发表的研究成果。由于本书按照 72 学时的要求进行编写，因此免疫遗传学、发育遗传学、细胞分化和癌症发生等重要内容就无法按专题讲授，只在有关章节作蜻蜓点水式的叙述。限于时间和水平,疏漏之处在所难免,欢迎读者批评指正。

作者 1990 年 3 月 12 日于

南京大学生物化学系

# 目　　录


第一章　遗传的物质基础——DNA …… 1
第一节　DNA 携带着两类不同的遗传信息 …… 1
第二节　DNA 的一级结构 …… 1
第三节　DNA 的二级结构 …… 3
一、Watson-Crick 右手双螺旋结构 …… 3
二、决定双螺旋结构状态的因素 …… 3
第四节　DNA 物理结构的不均一性 …… 7
一、反向重复序列 …… 7
二、富含 A/T 的序列 …… 7
三、嘌呤和嘧啶的排列顺序对双螺旋结构稳定性的影响 …… 8
第五节　DNA 双螺旋结构的呼吸作用 …… 9
第六节　DNA 的变性、复性、杂交和 Cot 曲线 …… 9
一、变性 …… 10
二、复性 …… 10
三、杂交 …… 11
四、Cot 曲线 …… 11
第七节　DNA 二级结构的多形性 …… 13
第八节　DNA 构象家族 …… 19
第九节　DNA 超螺旋和拓扑异构现象 …… 21
第十节　常见的 DNA 分子形式及其相互关系 …… 27
提要 …… 28
参考文献 …… 29
第二章　有机体、染色体和基因 …… 31
第一节　原核生物和真核生物 …… 31
第二节　基因组大小与 C 值矛盾 …… 32
第三节　原核生物染色体及其基因 …… 33
一、大肠杆菌染色体 …… 33
二、噬菌体 …… 36
第四节　真核生物的染色体 …… 39
一、真核生物 DNA 复性动力学 …… 39
二、真核生物染色体上的单一序列和重复序列以及卫星 DNA …… 41
三、卫星 DNA 的等级结构及其起源和进化 …… 42
四、染色质和核小体 …… 47
五、着丝点 …… 54
六、端粒 …… 55
第五节　真核生物的基因 …… 56

一、不连续基因……56
二、基因家族与基因簇……59
三、串联重复基因……61
四、细胞器基因……61
第六节 基因定位……67
一、遗传交换定位法……67
二、接合定位法……68
三、染色体步行和染色体跳跃……70
第七节 基因的分子进化……71
第八节 早期生命进化的三界系统理论……75
一、原核生物之间的巨大差异……75
二、基于 16S rRNA 碱基变化的通用系统发生分类法……76
三、内共生作用参与了真核生物的进化……77
提要……79
参考文献……80
第三章 DNA 的复制……84
第一节 DNA 的半保留复制……84
第二节 复制原点、方向和方式……84
第三节 DNA 复制的酶学……88
一、DNA 的聚合反应和聚合酶……88
二、脱氧核苷三磷酸前体的来源……89
三、三种 DNA 聚合酶的结构和功能……89
四、DNA 连接酶……93
五、与 DNA 几何学性质相关的酶……94
第四节 DNA 复制的半不连续性……95
一、DNA 半不连续性复制的发现……95
二、引物和引发酶……97
三、前体片断的连接……98
第五节 DNA 复制机构的复杂性……100
第六节 DNA 复制的起始……101
一、先导链合成的起始(上)：从新起始……101
二、先导链合成的起始(下)：共价延伸……109
三、后随链的前体片断的起始……112
四、由复制体进行先导链和后随链的同时复制……116
第七节 复制的终止……116
一、环形 DNA 复制的终止……116
二、线形 DNA 复制的终止……117
第八节 真核生物的 DNA 复制……120
一、真核生物的复制原点、复制元和复制元族……120
二、真核生物的 DNA 聚合酶和引发酶……120
三、SV40 的大 T 抗原与复制原点……122
四、SV40 以及其他真核生物的 DNA 复制过程……122

五、真核生物染色体末端 DNA 的复制……123
六、真核生物复制过程中的核小体结构……126
第九节　复制的调控……128
提要……131
参考文献……132
第四章　以修复作用为中心的 DNA 的安全保障体系……135
第一节　复制修复……135
一、尿嘧啶糖基酶系统……136
二、错配修复系统……137
第二节　损伤修复……138
一、胸腺嘧啶二聚体的产生……138
二、胸腺嘧啶二聚体修复的生物学指征……139
三、胸腺嘧啶二聚体修复的分子生物学机制……139
四、其他损伤类型及其修复……147
第三节　限制与修饰……150
一、限制-修饰现象……150
二、限制-修饰系统……151
三、限制-修饰系统的生物学意义……157
提要……157
参考文献……157
第五章　突变……159
第一节　概述：突变定义及其分类……159
第二节　条件型突变……161
第三节　回复突变和抑制突变……162
一、回复突变的鉴定和分类……162
二、基因内抑制突变……162
三、基因间抑制突变……164
四、基因间间接抑制突变……170
第四节　突变剂和突变生成……171
一、碱基类似物在 DNA 复制时的渗入……171
二、DNA 分子上碱基的化学修饰……173
三、嵌合剂的致突变作用……175
四、转座成分的致突变作用……176
五、增变基因……176
六、紫外线的致突变作用……176
七、突变热点……178
第五节　离体定向诱变……178
提要……182
参考文献……182
第六章　转录……184
第一节　概述……184

第二节 RNA 合成的酶学 …… 185
一、RNA 合成的基本特征 …… 185
二、*E.coli* RNA 聚合酶 …… 185
三、真核生物的 RNA 聚合酶 …… 187
第三节 控制转录起始的 DNA 序列——操作子和启动子的结构 …… 187
一、操纵元及其结构 …… 187
二、原核生物的启动子结构 …… 188
三、鉴别 RNA 聚合酶等蛋白质在 DNA 上结合位点的方法 …… 192
四、真核生物的启动子 …… 193
第四节 转录的起始和延伸 …… 197
第五节 转录的终止 …… 200
一、终止序列和释放因子 …… 200
二、终止与抗终止及抗终止因子 …… 202
三、真核生物转录的终止 …… 206
第六节 转录后处理(一) 转录产物的修饰——真核生物 mRNA 的帽子和尾巴 …… 207
一、帽子 …… 207
二、多聚(A)尾巴 …… 208
第七节 转录后处理(二) 基因间序列的去除——稳定 RNA 从多基因转录产物中的分离 …… 210
一、不稳定 RNA 与稳定 RNA …… 210
二、rRNA 的转录后处理 …… 211
三、tRNA 的转录后处理 …… 214
第八节 转录后处理(三) 内元的去除——RNA 的拼接 …… 216
一、概述 …… 216
二、tRNA 的拼接 …… 217
三、第Ⅰ类内元的拼接 …… 221
四、编码 RNA 成熟酶的内元 …… 228
五、第Ⅱ类内元的拼接 …… 231
六、核基因 mRNA 内元的拼接 …… 233
第九节 转录后处理(四) 不连续转录和反式拼接 …… 236
第十节 逆转录 …… 238
提要 …… 240
参考文献 …… 242
第七章 翻译 …… 247
第一节 tRNA 和遗传密码 …… 248
一、tRNA 的结构 …… 248
二、密码子与反密码子 …… 249
三、副密码子与氨酰基 tRNA 的合成 …… 250
四、tRNA 的丰富度与密码子的使用频率 …… 251
五、密码的通用性，线粒体密码的特殊性和密码的进化 …… 254
第二节 核糖体——制造蛋白质的工厂 …… 257
一、核糖体的结构 …… 257

二、核糖体的装配 …… 259
三、核糖体突变 …… 260
四、核糖体的活性位点 …… 260
第三节 肽链的合成 …… 262
一、合成的起始 …… 262
二、延伸 …… 265
三、终止和肽链的释放 …… 267
第四节 mRNA 的结构与翻译 …… 269
一、原核生物的多顺反子 mRNA 的翻译 …… 269
二、真核生物核糖体形成位点与双功能 mRNA …… 269
三、只有最后一个终止密码子的多基因 mRNA 的翻译 …… 270
第五节 蛋白质在细胞内的越膜运输，定位和翻译后处理 …… 271
一、概述 …… 271
二、细菌中蛋白质的越膜 …… 271
三、真核生物蛋白质的转运、分拣和锚定 …… 271
四、翻译后处理 …… 276
提要 …… 277
参考文献 …… 278
第八章 原核生物基因表达的调控 …… 281
第一节 概述 …… 281
第二节 正调控与负调控 …… 281
第三节 操纵元的原型——乳糖操纵元 …… 283
一、操纵元模型的提出 …… 283
二、乳糖操纵元的调控机理 …… 285
三、阻遏蛋白与操作子的相互作用 …… 286
四、乳糖操纵元的正调控 …… 290
第四节 操纵元的其他调控形式 …… 291
一、具有双启动子的半乳糖操纵元 …… 291
二、阿拉伯糖操纵元：具有双重功能(正控制和负控制)的调节蛋白质 …… 293
三、色氨酸操纵元：可阻遏系统 …… 296
第五节 基因转录的时序调控 …… 298
一、枯草杆菌中 σ 亚基的更迭 …… 298
二、大肠杆菌热震惊基因的表达 …… 301
三、T4 噬菌体生长中 RNA 聚合酶亚基的修饰和 σ 亚基的替换 …… 302
四、T7 噬菌体生长过程中用噬菌体 RNA 聚合酶代替寄主的 RNA 聚合酶 …… 303
五、λ 噬菌体早期、晚期基因转录以及裂解生长和溶原生长的调控 …… 306
第六节 基因转录的翻译调控——衰减子系统 …… 310
一、概述 …… 310
二、衰减子的作用机制 …… 310
三、衰减子的普遍性及其生物学意义 …… 313
四、负责核苷酸合成的基因的衰减子 …… 314
第七节 翻译水平的调控 …… 316

一、反义 RNA 的调控作用 …… 316
二、mRNA 本身的二级结构影响翻译的进行 …… 317
三、mRNA 的寿命对基因表达的调控 …… 319
四、蛋白质合成的自体调控 …… 321
五、严谨反应 …… 325
第八节 DNA 序列重排对基因转录的调控 …… 326
提要 …… 328
参考文献 …… 329
**第九章 真核生物基因表达的调控** …… 332
第一节 概述 …… 332
一、真核生物基因调控的特点 …… 332
二、活跃表达基因的数目 …… 332
三、基因表达的不同水平，丰富 mRNA 和稀少mRNA …… 333
四、持家基因和奢侈基因 …… 335
五、胚胎分化过程中有可能存在着相对简单的分子开关 …… 336
第二节 DNA 水平的调控 …… 337
一、基因丢失 …… 337
二、基因扩增 …… 338
三、基因重排 …… 339
第三节 活跃转录基因的染色质结构 …… 341
一、转录基因的核小体结构 …… 341
二、DNA 酶 I 优先敏感性和 HMG 蛋白质 …… 342
三、DNA 酶 I 超敏感点 …… 343
四、组蛋白的修饰作用 …… 345
五、DNA 的甲基化和去甲基化 …… 346
第四节 转录水平的调控 …… 348
一、Britten-Davidson 模型 …… 348
二、基因调控的顺式作用成分 …… 350
三、基因调控的反式作用因子 …… 354
四、真核基因转录调控的机制 …… 356
第五节 rRNA 基因的转录：先终止还是先起始？ …… 359
第六节 意外的意外，RNA 聚合酶Ⅲ启动子的多形性和三种 RNA 聚合酶及其启动子的共通性 …… 360
第七节 真核生物中转录的翻译控制 …… 361
第八节 转录后水平的调控 …… 363
一、hnRNA 的选择性加工运输 …… 363
二、mRNA 前体的选择性拼接 …… 364
第九节 翻译水平的调控 …… 368
一、mRNA 的稳定性 …… 368
二、mRNA 翻译起始的调控 …… 369
三、真核生物蛋白质合成的自体调控 …… 370

提要……371
参考文献……372
第十章 基因表达调控的分子机制……377
第一节 概述……377
第二节 两类遗传信息的差异……377
第三节 调节蛋白质的对称性与其识别序列的对称性……379
第四节 λ噬菌体的阻遏蛋白和 Cro 蛋白所组成的转录开关系统……380
第五节 调控蛋白质的结构花式及其与 DNA 相互作用的细微机制……386
一、概述……386
二、α 螺旋—转角—α 螺旋……387
三、锌指结构……394
四、Leu 拉链……395
第六节 不同结合位点上的蛋白质之间的相互作用……396
一、蛋白质相互作用的方式……396
二、蛋白质相互作用的复杂性和正、负调控作用的相对性……398
第七节 RNA 开关系统……400
提要……401
参考文献……402
第十一章 遗传重组……404
第一节 概述……404
第二节 同源重组的分子机制……405
一、断裂—复合及 Holliday 中间体的形成……405
二、Holliday 中间体的拆分……408
三、异源双链与基因转换……411
四、细菌转化、接合和转导的重组机制……413
第三节 同源重组的酶学机制——RecA 蛋白质和 RecBCD 蛋白质在同源重组中的作用……415
一、RecA 蛋白质在联会和链交换中的作用……415
二、RecA 蛋白质和 Holliday 中间体的形成……419
三、RecBCD 在同源重组中的作用和重组热点……420
四、拆分 Holliday 中间体的酶活性……423
五、RecF 途径……423
第四节 依赖于同源重组的位点特异性的序列代换——酵母 MAT 序列的转换……424
第五节 位点特异性重组……424
一、λ噬菌体 DNA 的整合与切除……424
二、λ噬菌体整合的分子机制……427
第六节 细菌中的转座成分……429
一、转座成分概述……429
二、插入序列……430
三、复合转座元……431
四、Tn10 和 Tn5 的末端组件与转座的调节……434

五、转座的机制 …… 436
六、Tn3 及其转座作用 …… 436
七、作为转座成分的 Mu 噬菌体 …… 441
八、沙门氏菌鞭毛相变中的转座机制 …… 443
第七节 真核生物中的转座成分(一) 真核生物转座成分的分类及其典型代表 …… 443
一、真核生物中的转座成分及其分类 …… 443
二、酵母的 Ty 成分 …… 445
三、果蝇的 copia 等转座成分 …… 446
四、玉米的转座成分 …… 449
第八节 真核生物中的转座成分(二) 还原病毒 …… 450
一、通过 DNA 中间体复制其基因组 RNA …… 450
二、还原病毒的基因组结构 …… 451
三、还原病毒基因组 RNA 到原病毒 DNA 的转变过程 …… 451
四、原病毒的基因表达 …… 456
五、还原病毒的癌基因及其起源 …… 456
六、还原病毒和转座元的关系 …… 459
第九节 真核生物中的转座成分(三) 非病毒返座元 …… 460
一、概述 …… 460
二、RNA 聚合酶Ⅱ转录产物的返座假基因和返座基因 …… 461
三、RNA 聚合酶Ⅲ转录产物的返座元 …… 461
四、关于返座的机制 …… 463
提要 …… 464
参考文献 …… 466
第十二章 基因工程导论 …… 469
一、概述 …… 469
二、限制性内切酶对 DNA 分子的专一性切割 …… 469
三、载体：克隆载体，穿梭载体和表达载体 …… 471
四、基因组克隆和 cDNA 克隆 …… 475
五、核苷酸链的杂交反应提供了检测特定核苷酸片断的灵敏方法 …… 478
六、用 DNA 重组技术大量制造细胞中的微量蛋白质 …… 479
七、植物的基因工程 …… 480
八、基因工程工业 …… 482
九、结束语 …… 486
参考文献 …… 486
附录：涵义和译名有所变化的部份分子遗传学术语 …… 488
索引 …… 493

# 第一章　遗传的物质基础——DNA

## 第一节　DNA 携带着两类不同的遗传信息[1,2]

除了少数的 RNA 病毒之外，DNA 几乎是所有生物遗传信息的携带者。DNA 分子携带着两类不同的遗传信息。一类是负责蛋白质氨基酸组成的信息，是以众所周知的三联体密码子方式进行编码的。在 64 个三联体密码子中除了三个终止密码子外，其余 61 个密码子代表了 20 种不同的氨基酸，因而大多数氨基酸都有一种以上的密码子，这就是所谓密码子的简并性。几乎所有的生物都使用相同的密码子。这就是所谓密码子的通用性。后来发现，线粒体的密码子与核基因的密码子不同，这可能与线粒体的蛋白质合成机器有关。人们还发现，密码子的碱基组成与氨基酸的侧链性质有一定的相关性。近年来人们又发现了硒半胱氨酸的密码子以及密码子进化方面的规律。此外，人们还发现了决定 tRNA 本质的副密码子(paracodon)。

DNA 分子上所携带的另一类遗传信息是关于基因选择性表达的信息。我们知道，在原核生物中，结构基因占基因组的比例很大。特别是在亚细胞生命体噬菌体中，这一比例就更高。例如 $\phi \times 174$ 噬菌体共有碱基 5 386 个，结构基因用去 5 169 个，这个比例高达 96%。然而，在高等哺乳动物中，这一比例只有大约 10～15%。人们不禁要问，其余占基因组 80% 以上的DNA 起什么作用？目前对这个问题还无法作出精确解释；但可以肯定的一点是，基因组中大部分 DNA 序列是用来编码基因选择性表达的遗传信息的。这种选择性表现在，细胞周期的不同时相中，个体发育的不同阶段中，不同的器官和组织中，在不同的外界环境下，各种基因是关闭还是表达，表达量是多少，都是各不相同的。用来编码基因选择性表达的遗传信息的这一部分 DNA 序列并不是单独发挥其调控作用的，在这一点上它与结构基因 DNA 序列是相类似的。然而，二者还是有差别的。结构基因的转录和翻译，都是强烈地依赖于转录酶系和一整套蛋白质合成机构，而核酸序列本身的差异只是提供了这些酶和蛋白质识别的基础。而对于调控 DNA 序列来说，除了上面作用之外，可能还会主动修饰自己的双螺旋空间结构。说得更确切些，调控 DNA 序列能与调控蛋白质相互作用而修饰自己的双螺旋空间结构，以便更好地为调控蛋白质所识别。美国 E．W．Prohofsky 借助于固体物理学技术，非常成功地预言了DNA的大频率跨度(从无线电波到红外)的电磁波谱的振动模。与生物学和基因表达有关的大部分信息存在于长程力所引起的低频振动，使得 DNA 的各部分序列之间能进行长距离的对话，从一端可以了解另一端发生了什么情况。

## 第二节　DNA 的一级结构[3,4]

DNA 从结构上来说是由脱氧核糖核苷单磷酸通过 3′，5′ 磷酸二酯键连接而成的高聚物。从同一个磷酸基的 3′ 酯键到 5′ 酯键的方向定为链的方向。在大多数天然DNA分子长链的两端，总是有一个核糖带有自由的 5′- 磷酸，而另一端的核糖带有自由的3′-羟基，前者称为5′端，后者称为 3′ 端。DNA 链的方向就是从 5′ 端到 3′ 端。习惯写法中，如 pApCpApGpT，左边总

是 5′ 端，右边是 3′ 端，OH 一般省略不写，这已成常规。

DNA 分子一级结构中的显著特点，顾名思义，就是脱氧核糖了。也就是核糖的 2′ 位上没有自由羟基，这也就是 DNA 这一主要遗传物质极其稳定的根本原因。特别是对于碱的抵抗力，在 pH11.5 时，DNA 链的一级结构几乎没有任何变化，而 RNA 链在几分钟内降解为 2′-单磷酸核苷和 3′-单磷酸核苷。RNA 碱水解先生成一个 2′,3′ 环式单核苷酸中间产物，由于是一个五员环，稳定性较差，很快转变成 2′ 单核苷酸和 3′ 单核苷酸，见图 1-1。

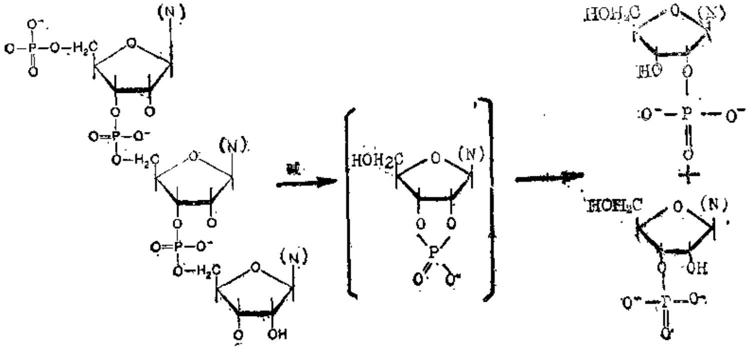



图 1-1 RNA 由于含有 2′-OH，很容易被碱水解

DNA 一级结构的测定即 DNA 的序列分析工作在近十多年中进展迅速，Sanger 在这方面作出了最为杰出的贡献。众所周知，Sanger 于 60 年代初利用小片段重迭法首次测定了一个生物大分子胰岛素的 51 个氨基酸序列。此后不少人使用小片段重迭法测定 DNA 和 RNA 的核苷酸序列，结果只测定出若干小分子的 RNA(如 $tRNA^{Ala}$)；而 DNA 序列分析仍然步履艰难。因为 DNA 分子巨大，而组成这种巨大分子的单体又只有四种，因此从战略上来说，用小片段重迭法去进行 DNA 的序列分析是不恰当的。1975 年，Sanger 提出了新的战略，也就是用测定 DNA 片段的相对长度来推断核苷酸顺序的方法。在 8～20% 的聚丙烯酰胺凝胶上电泳 40 厘米左右，即可区分一个核苷酸之差的 10～160 个核苷酸的各种大小的片段。每一种片段含量很低，必须用 $^{32}P$ 进行标记，通过放射自显影才能得到清晰的图象。这就是所谓"凝胶直读法"，即以凝胶电泳为主要手段，从放射自显影的 X-光胶片上直接读出核苷酸序列。凝胶直读法又可分为两类：一类是酶法，Sanger 于 1975 年设计的"加减法"以及他们于 1977 年所改进的"末端终止法"（即"双脱氧法"）都是酶法。这种方法要求有单链 DNA 模板和与之互补的一段引物，然后以 DNA 聚合酶 I 进行互补链的合成。因此从凝胶电泳的放射自显影 X-胶片上读出的序列是互补链的序列。这种方法又称为间接拷贝法（indirect copying methods）。其中双脱氧法操作简便、迅速，已成为 DNA 序列分析的首选方法。另一类是化学法，即 1977 年 Maxam 和 Gilbert 所设计的化学断裂法，针对四种核苷酸使用不同的链断裂方法。这种方法也简便、精确，对于较短的 DNA 片段分析起来比双脱氧法快得多，因此目前也被广泛采用。根据 Sanger 提出的战略，人们还发展了许多测定 RNA 序列的方法。对于序列分析的具体方法，本书不能作详细介绍，请读者参阅有关文献。

DNA 一级结构的重要意义不仅在于它蕴藏了遗传信息（以密码子的方式），而且还决定

了 DNA 的二级结构和空间结构。特别是1977年 Sanger 首次测定了噬菌体 $\phi\times174$ DNA 的一级结构，不但揭示了该噬菌体的基因组织情况和基因重叠现象，而且也为研究 DNA 的二级结构和空间结构奠定了坚实的基础。从结构与遗传学功能的关系上来看，从一级结构中得到的许多信息都需要用二级结构和空间结构来进行解释；还有大量的遗传特征需要从二级结构和空间结构上加以发掘。二级结构和空间结构必将成为弄清生命遗传奥秘的主要研究领域之一。

## 第三节　DNA 的二级结构[3~6]

### 一、Watson-Crick 右手双螺旋结构

Watson 和 Crick 在 1953 年 4 月提出了 DNA 的右手双螺旋结构，绝不是偶然的，而是许多科学家艰苦细致，甚至一时看来可能是毫无意义的工作结晶。早在 40 年代，W.T.Astbury 就用 X-光衍射研究DNA 的结构；1950 年，Chargaff 从大量的不同来源的 DNA 样品的分析中发现了 DNA 组成的当量规律，即 A=T,G=C,A+G=T+C；1952 年M.H.Wilkins 用高度定向的 DNA 纤维作出了极其出色的 X-光衍射照片；Watson 和 Crick 早就设想过 DNA 结构的多种可能的模型，但它们缺乏高质量的 X-光衍射照片；当他们得到 Wilkins 的高质量的 X-光 衍射照片时，很快就发表了他们的右手双螺旋结构模型。

右手双螺旋结构模型的要点：

1. *主链*　脱氧核糖和磷酸基通过 3′,5′ 磷酸二酯键交互连接，成为螺旋链的骨架。二条主链以反向平行的方式组成双螺旋，螺旋的直径为 20Å 左右。主链处于螺旋的外侧，这是由核糖和磷酸的亲水性所决定。核糖平面与螺旋轴平行。碱基则处于螺旋的内侧。

2. *碱基对*　由于双螺旋结构要求有一个正常的螺旋形式，就必须是嘌呤和嘧啶相配。否则两个嘌呤尺寸太大，超过了 20Å；而两个嘧啶的尺寸太小，又不足 20Å。而 A 和 C,G 和 T 又不能形成适合的氢键，因此只有 A 和 T 相配，G 和 C 相配才能满足正常螺旋形式的要求，而且也只有这样才能满足 Chargaff 的当量的规律。这样的碱基配对规律，叫做碱基互补。碱基平面与螺旋轴基本上是垂直的。嘌呤环和嘧啶环上的氨基和酮基是亲水的，因而配对碱基之间能够形成氢键；而嘌呤环和嘧啶环本身却是疏水的；因而同一条链中的相邻碱基能够形成一种堆积力。这两种力的协同作用就维持了双螺旋结构的稳定性。

3. *螺距*　双螺旋链中的任意一条链绕轴一周 (360°) 所升降的距离的叫做螺距。这个模型的螺距为 34Å，其中包含 10 个核苷酸，因此每两个相邻碱基平面的垂直距离是 3.4Å，相对于螺旋轴移动 36°。

4. *大沟和小沟*　沿螺旋轴方向观察，两条主链和碱基并不充满双螺旋的空间，双螺旋的表面形成两条凹槽，一条宽而深，叫做大沟，一条狭而浅，叫做小沟。之所以形成大沟和小沟，是因为从螺旋轴心到两条主链所划分出的两个扇形不等，一个大于 180°，一个小于 180°。这两条沟，特别是大沟，对于在遗传上有重要功能的蛋白质识别 DNA 双螺旋结构上的特定信息是非常重要的，因为只有在沟内，蛋白质才能"感觉"到不同碱基顺序，而在双螺旋结构的表面全是相同的磷酸和脱氧核糖的骨架，是没有什么信息可言的。

### 二、决定双螺旋结构状态的因素

1. *氢键*　在每一个碱基上都有适于形成氢键的供氢体如氨基和羟基以及受氢体如酮基

和亚氨基。G·C 对之间有三条氢键，A·T 对之间只有两条氢键。这是 DNA 双螺旋结构的重要特征之一，DNA 的许多性质都与此有关。

双螺旋结构中存在氢键，可以用 DNA 的熔解温度 $T_m$ 与 G+C 的百分含量成正比来得到说明。在 G+C 百分含量在 30% 到 70% 的范围内，在 0.15mol/L NaCl+0.015mol/L 柠檬酸钠溶液中，$T_m$ 与碱基组成的关系可以用下面公式表示：

$$T_m = 69.3 + 0.41(\%G + C)$$

这就是所谓 Marmur-Doty 关系式，图 1-2 即表示这种关系。如果溶液中加入尿素或甲酰胺，则可使 $T_m$ 显著降低。因为正常情况下，DNA 双螺旋结构中的氢键处于不断的断裂和复合的热平衡状态，而这两种试剂能与碱基形成氢键而减少碱基与碱基形成氢键的机会。

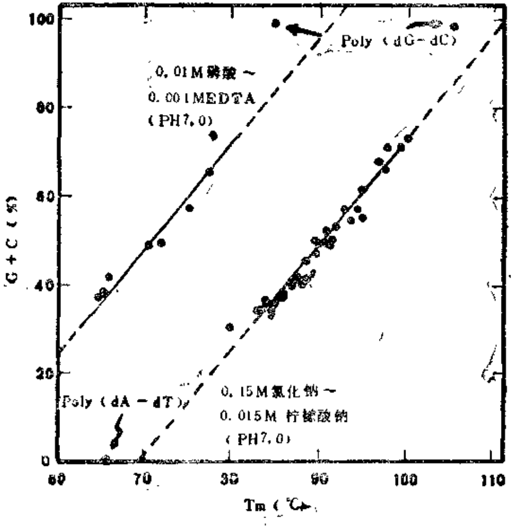



图 1-2 (G+C)含量与 $T_m$ 的关系

在形成氢键时，要求供体氢原子与受体氢原子具有互补性，因而特异性很强。受体原子到供体原子的距离称为氢键的键长。DNA互补碱基之间的氢键键长2.8Å～3.0Å，键能4～6千卡/mole。氢键的另一个特点是具有高度的方向性。如果供体原子、氢原子和受体原子三者在一条直线上，形成的氢键最强。同链的相邻碱基之间的堆积力恰恰为形成最强的氢键提供了条件。

2. *碱基堆集力* 同一条链中的相邻碱基之间的非特异性作用力，即疏水作用力，和 van der Waal 力。

疏水作用是不溶于水或难溶于水两个分子(相同或不同种类)在水中具有相互联合，成串地结合在一起的趋势，其中并无键的生成，但这样的趋势在热力学上是有利的。这可以简略地解释如下。水分子能相互形成氢键。如果一种分子能与水形成氢键，则能溶于水。然而这种分子与水之间的氢键并没有水分子之间氢键那样牢固。因此，即使是极性最强的物质，其溶解度一般也是有限的。当把不溶于水的(即不能与水形成氢键的)有机分子放入水中，这些分子

靠 van der Waal 力彼此附着。实际上在水中很难插入不能形成氢键的有机分子。然而 DNA 是一大分子的高聚物，其主链部分是亲水的，即使是碱基，其外围基团如氨基、酮基也是亲水的，并能与水分子形成氢键。然而，嘌呤环和嘧啶环本身却带有一定程度的疏水性。一个疏水的分子插入水中，必然引起一个有序的以氢键相连的水分子层绕着，这就意味着熵值的减少，从热力学第二定律来看不可能是一个自发的过程。如果这些不溶性的环状物相互联合，则可以使有序的水分子壳层减少到很低的数量，使得所有的分子(水和溶质)的熵值更大一些，因此，DNA 相邻碱基就有相互堆集在一起的趋势，这是形成碱基堆集力的重要因素之一。

另一方面，由于双螺旋结构中相邻碱基垂直距离为 3.4Å，而嘌呤环和嘧啶环的 van der Waal 半径恰好为 1.7Å 左右，这样，van der Waal 力加强了疏水作用力。van der Waal 作用力大约 1 千卡/mole。根据能量的玻茨曼分布，在室温下分子热运动的平均能量是 0.6 千卡/mole，因此两个原子间的 van der Waal 力是没有什么实际意义的。然而，若干对原子，特别是 DNA 链中的大量的嘌呤环和嘧啶环，其累积的 van der Waal 力是相当可观的，是构成碱基堆积力的另一个重要因素。

DNA 双螺旋结构中存在着疏水作用力，可以用加入某些能增加碱基溶解度的试剂(如甲醇)或破坏碱基周围有序的水壳层的试剂(如三氟乙酸钠)得到证明。这些试剂对于磷酸–脱氧核糖骨架均无明显影响，然而它们能显著地降低 $T_m$ 值，$T_m$ 值降低的程度恰恰与碱基的溶解度的增加成比例。许多结果都表明，多核苷酸倾向于有一个三维结构以使高度溶解的磷酸基团与水保持最大的接触而把碱基与水的接触减少到最低限度。这就解释了为什么在 DNA 中磷酸脱氧核糖链在外面而碱基在里面。

用旋光色散(ORD)和圆二色性(CD)对单链 DNA 和 RNA 进行研究的结果表明，在单股多核苷酸中大部分碱基平行排列使长链成为有规则的螺线 (helix) 而不是无规则线团 (random coil)。所谓无规则线团就是：在一个高聚物的长链上，没有任何链内相互作用，每一单体可以相对于相邻的单体自由旋转，其限制仅仅在于两个原子不能占据同一个空间。一条链的这种三维空间构型就叫做无规则线团。见图 1–3。

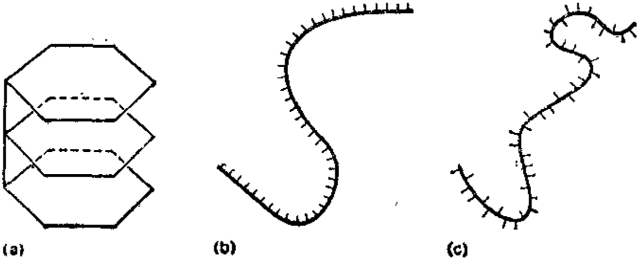



图 1–3　碱基平行排列(a)、规则螺线(b)和无规则线团(c)

人们甚至在二核苷酸中也发现了碱基平行排列的趋势，在 DNA 中则更是如此。这些都表明碱基堆积作用的存在。关于 DNA 的碱基堆集作用，人们还发现如下事实：

(1) 加入能减弱疏水作用的试剂可以消除碱基堆集作用。

(2) DNA 样品加热时，碱基堆集作用减少，同时伴随 $A_{260}$ 的增加。

(3) 能够断裂氢键的试剂对于单链 DNA 的碱基堆集作用没有影响，但是将双链 DNA 碱基堆集作用减少到变性 DNA 的程度。

从以上事实可以得出结论：双链 DNA 中的碱基比单链 DNA 中的碱基堆集程度更高，这是由两条链的配对碱基之间的氢键所引起的。氢键和疏水作用力都易被热运动所减弱和破坏。当所有的碱基指向正确的方向时，可以达到最大的氢键键合；同样，如果碱基不能歪斜或摆动，则能提高碱基堆集作用。很明显，已经堆集的碱基更容易氢键键合；相应地，已被氢键定向的碱基更容易堆集。因此，两种作用力相互协同，形成一种非常稳定的结构。如果一种作用力被消除，则另一种作用力大为削弱。这可以解释为什么加入破坏其中一种作用力的试剂会使 $T_m$ 值急剧降落。

3. 带负电荷的磷酸基的静电斥力　每一个核苷酸的磷酸基上都带有一个负电荷。如果这些负电荷没有被中和，双链之间的这种强有力的静电排斥作用将驱使两条链分开（在同一条链内虽然也存在着这种静电斥力，但由于链内的共价键，这种静电斥力并不重要）。但是当有盐类加入时，这些带负电荷的磷酸基团可以被正离子（如 $Na^+$）所中和，也就是正离子围绕在磷酸基周围形成了"离子云"，有效地屏蔽了磷酸基之间的静电斥力。这就是 Debye-Hückel 离子屏蔽理论。在生理盐浓度(约 0.2mol/L)时就发生了这种屏蔽作用，斥力也被中止。当离子浓度降低时，这种屏蔽作用减弱，斥力增大，因而 $T_m$ 值随之降低。事实上，纯蒸馏水中的 DNA 在室温下就会变性，也就是这个缘故。在超过生理盐浓度时，$T_m$ 值仍然随着离子强度的增加而上升，这是因为在高盐浓度下碱基的溶解性降低而增加了疏水作用力，促进了双螺旋结构的稳定。$T_m$ 值与盐浓度的关系示于图 1-4。

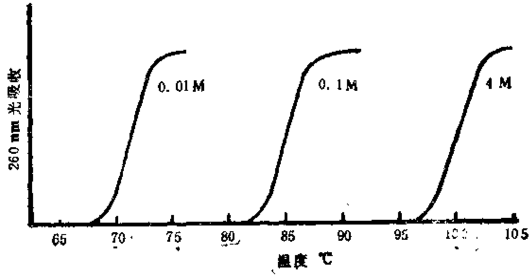


图 1-4　$T_m$ 值与中性盐 NaCl 浓度的关系

事实上，平时制备的 DNA 也总是 DNA 的钠盐，每个磷酸基总是结合一个 $Na^+$。在用 NaCl 和 CsCl 测量 DNA 分子量时，就会得到不同的数值，二者之比为 0.75。因为单核苷酸的平均分子量为 330，若是钠盐则为 353，若是铯盐则为 467。353/467 = 0.75。这就是一个有力的证据。

4. 碱基分子内能　当由于温度等因素使碱基分子内能增加时，碱基的定向排列遭受破坏，从而削弱碱基的氢键结合力和碱基的堆集力，会使 DNA 双螺旋结构受到破坏。

由此可见，在决定 DNA 双螺旋结构状态的四点因素中，前二者（互补碱基的氢键结合力和相邻碱基的堆集力）有利于 DNA 维持双螺旋构型，而后二者(磷酸基的静电斥力和碱基分子的内能）则不利于 DNA 维持双螺旋构型。一种 DNA 分子的结构状态将是这四种因素竞争的结果。

## 第四节 DNA 物理结构的不均一性(heterogeneity)[7~9]

在 DNA 的一级结构中，四种碱基 A、T、G、C 远非均匀分布，尽管双螺旋的构型大体相同，但沿着 DNA 长链各处的物理结构并不完全相同，各处双螺旋的稳定性有所差异，有时有显著的差异，甚至于在一定条件下改变了双螺旋的构型，如第七节将要讨论的 Z 型 DNA。这些都体现了大分子一级结构决定高级结构的原理。除了 Z 型 DNA 外，现介绍如下不均一性。

### 一、反向重复序列(inverted repeats)

反向重复序列又称为迴文序列(Palindrome)，它能在 DNA 或 RNA 中形成发夹结构，在 DNA 中有可能形成十字形结构。这一转变过程(见图1-5)必须伴随着一定的能量输入，因此一般认为在DNA水平上发生这种转变的可能性较小。

较短的迴文结构可能是作为一种特别信号，如限制性内切核酸酶的识别位点。较长的迴文序列容易转变为发夹结构，其功能目前还不完全清楚。但人们已经发现，转录的终止作用与迴文结构有关(详见转录一章)。发夹环在 tRNA 的结构中也很重要，每一种 tRNA 都是由几个发夹环组成的三叶草形结构。

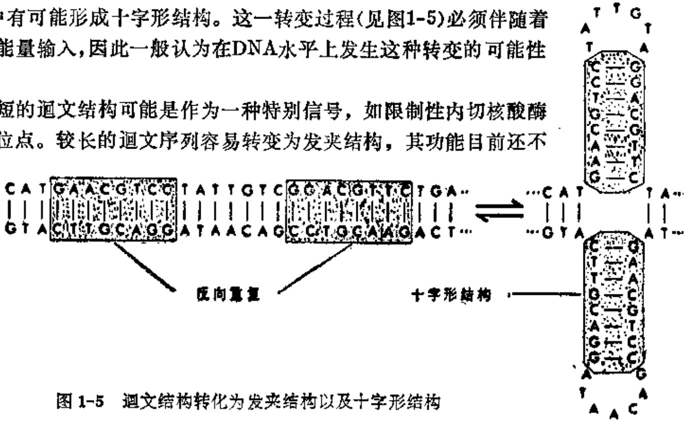



图 1-5 迴文结构转化为发夹结构以及十字形结构

### 二、富含 A/T 的序列

在高等生物中，A+T 与 G+C 的含量差不多相等。然而在它们的染色体的某一区域，A·T 含量可能高得惊人。例如螃蟹的卫星 DNA 区段，A·T 含量竟然高达 97%。它们仅仅重复 AT 两个碱基，即 ATATAT……；大约 30 个这样碱基才有时插入一个 G 或 C。这种卫星 DNA 的功能将在下一章中讲述。在很多有重要调节功能(而不是蛋白质编码)的 DNA 区段都富含 A·T，特别是在复制起点和启动子的 Pribnow 框（真核生物为 TATA 框）的序列中，都富含 A·T 碱基对，对于复制和转录的起始十分重要。因为 G·C 对含有三条氢键，而 A·T 对只有二条氢键，此处的双链容易解开，更有某些蛋白质促进这种过程，以利于起始复合物的形成(详见有关章节)。

在 DNA 分子中，由于特异性序列的存在，能使局部区域产生特殊的构象。例如在动基体 DNA 中出现的 CAAAAAT 或 CAAAAAAT 的重复序列，分布周期约为 10 个碱基对（一圈螺旋）。它们具有很高的电泳迁移率。这可能是由局部 DNA 形成扭结 (kinking) 所致。由

于扭结的存在,局部 DNA 发生部分解缠和弯曲。

## 三、嘌呤和嘧啶的排列顺序对双螺旋结构稳定性的影响

人们考察了十种相邻的二核苷酸对(nearest-neighbor doublets),得出一个非常有趣的现象,那就是碱基组成相同,但嘌呤和嘧啶的排列顺序不同,双螺旋的稳定性具有显著的差异。例如$\frac{5'GC3'}{3'CG5'}$和$\frac{5'CG3'}{3'GC5'}$的稳定性相差很大,前者的稳定性远大于后者。它们的氢键数目是相同的,它们的差别在于相邻碱基之间的堆集力不同。即从嘌呤到嘧啶的方向的碱基堆集作用显著地大于同样组成的嘧啶到嘌呤方向的碱基堆集作用。(这里的方向就是常规的从 5′ 端到 3′ 端的方向)。这是因为前者的嘌呤环和嘧啶环重迭面积大于后者的嘧啶环和嘌呤环的重迭面积,这在 B 型 DNA 中确是如此。

根据 Gotoh 1981 年的研究,十种相邻二核苷酸对的 $T_m$ 值如表 1-1 所示,单位为 ℃,所用离子强度为 19.5mmol/L $Na^+$。

表 1-1 相邻二核苷酸对 $T_m$ 值

| 5′ \ 3′ | A | T | G | C |
|---|---|---|---|---|
| A | 54.50 | 57.02 | 58.42 | 97.73 |
| T | 36.73 | 54.50 | 54.71 | 36.44 |
| G | 86.44 | 97.73 | 85.97 | 136.12 |
| C | 54.71 | 58.42 | 72.55 | 85.97 |

将 Gotoh 的 $T_m$ 值与 Ornstein(1978) 的碱基堆集能数值结合起来绘制相关图,相关系数接近于 1,见图 1-6。

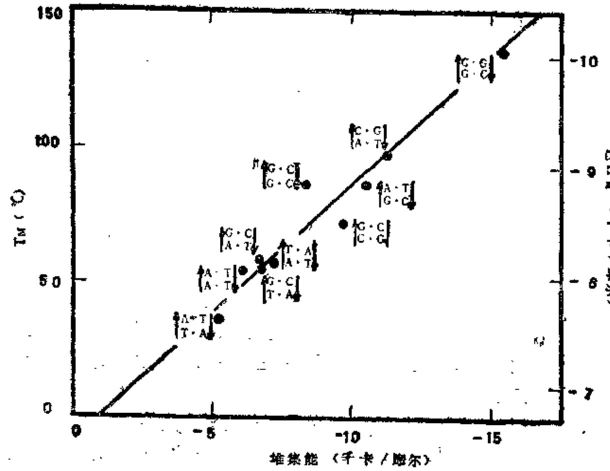



图 1-6 二核苷酸对的 $T_m$ 值与堆集能相关图

由表 1-1 可以看到，$\begin{matrix}\overrightarrow{TA}\\ \overleftarrow{AT}\end{matrix}$ 的 $T_m$ 值最低。在真核生物中，常可以在 −19 到 −27 的位置上看到一个叫做 TATA 框的结构(又称 Hogness 框)，这是 RNA 聚合酶的结合位点。在这里 RNA 聚合酶和有关蛋白质因子形成转录起始复合物。

又如，生命有机体选择 UAA 作为最有效的终止密码子绝不是偶然的，因为在 64 个三联体密码子中，它与反密码子(假定有的话)形成的互补产物$\begin{matrix}\overrightarrow{UAA}\\ \overleftarrow{AUU}\end{matrix}$的 $T_m$ 值是最低的一个，即使在生理温度下也是不稳定的。当初有人花了很多工夫去寻找一个不携带氨基酸的专供肽链终止用的 tRNA，其实并不存在这种 tRNA。肽链的释放是由释放因子 RF 在起作用。在三种终止密码子中，UAG 和 UGA 常会为突变型的 tRNA 无义抑制，而 UAA 则很少发生无义抑制也可能就是这个道理。这也就说明了为什么在肽链终止处常常会出现双重终止密码子。

由于相邻二核苷酸对稳定性的分析，使得计算机模拟的DNA熔解曲线更接近于实验所得出的数值。但由于生物体系的复杂性，在不同种的 DNA 之间，在不同大小的 DNA 片断之间，也常会出现某些差异，有些参数也还有待于进一步修正。

## 第五节　DNA 双螺旋结构的呼吸作用[10]

在 DNA 双螺旋结构中，配对碱基之间的氢键处于连续不断的断裂和再生的动态平衡之中，这可以通过甲醛变性试验得到证明。在 DNA 中各碱基上的氨基总是形成氢键的，而甲醛只能和自由氨基起反应。实际上，甲醛确实能和 DNA 的氨基起反应使 DNA 发生不可逆变性，就是因为当碱基之间氢键断裂时，甲醛才有机会与氨基起反应。由于断裂的氢键又可能立即恢复，甲醛要与全部氨基起反应需经过相当长的时间，如图 1-7 所示。

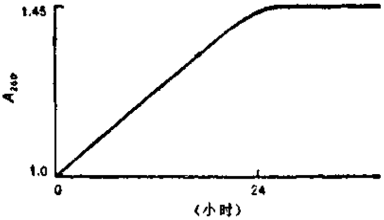


图 1-7　含有 4% 甲醛的 DNA 溶液的变性图(温度为25°C)

配对碱基之间的氢键不但处于断裂和再生的平衡状态中，氢键上的氢原子还能和水发生交换，这很容易用重水 ($^{2}H_2O$) 作溶剂得到证明。这种现象进一步证明了氢键的迅速地断裂和再生过程。这种过程通常称为 DNA 链的呼吸作用。特别是在富含 A·T 的节段，呼吸作用更为明显，经常发生瞬间的单链泡状结构，这对于某些特殊的蛋白质与 DNA 发生反应并阅读 DNA 链内部储藏的信息(包括氨基酸编码)具有重要作用。

在双链 DNA 的两端，有 3~7 个碱基对经常不同程度地处于单链状态，这种现象叫做绽裂(fraying)。某些具有单链特异性的脱氧核糖核酸酶，如 $BAL_{31}$，同时又具有双链核酸外切酶的活性，大概就是由双链末端的绽裂作用造成的。

## 第六节　DNA 的变性、复性、杂交和 Cot 曲线[11,12,13]

大多数天然存在的 DNA 都具有规则的双螺旋结构，当 DNA 被加热时或某些试剂的作

用下，配对碱基之间氢键和相邻碱基之间的堆集力就会受到破坏，逐步变为近似于无规则的线团构型即变性 DNA。由天然状态到变性状态的转变过程叫做变性，又称为熔解。能使DNA变性的试剂叫做变性剂，如尿素、甲酰胺等。变性的 DNA 为单股 DNA。变性 DNA 在一定条件下又可以恢复天然 DNA 的结构，这个过程叫做复性，也叫重退火 (reannealing)。前几节所述的关于 DNA 的结构和该结构内各种作用力大多是通过研究 DNA 变性和复性过程而获得的。

## 一、变　性

研究 DNA 变性过程的方法很多，导致变性的方法主要是加热变性和用变性剂处理。测量变性的方法很多，有分光光度法，流体力学法，量热法，极谱法，旋光色散法，圆二色性法，层析法等。使用最为广泛的方法是光吸收法(260nm)，然而圆二色性法也是非常有希望的方法，它能提供足够的信息以补充光吸收法的不足之处。

由于 DNA 中的碱基的嘌呤环和嘧啶环中 $\pi\sim\pi^*$ 的转变，在 260nm 有明显的吸收峰。当 DNA 变性时，有序的碱基排列被打乱了，于是增加了对光的吸收，记住下列数据是有用的：当浓度都为 $50\mu g/ml$ 时，双螺旋 DNA 的 $A_{260}=1.00$，完全变性 DNA 即单链 DNA的 $A_{260}=1.37$，单核苷酸的等比例混合物的 $A_{260}=1.60$。由于 DNA 变性而引起的光吸收的增加称为增色效应(hyperchromicity)。

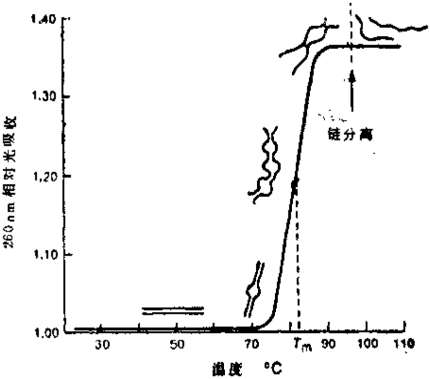


图 1-8　DNA 的熔解曲线以及各段温度范围内 DNA 可能采取的构型

当缓慢而均匀地(目前可以做到0.1℃/分）增加 DNA 溶液的温度，记录各个不同温度下的 $A_{260}$ 数值，即可绘制成 DNA 的熔解曲线，如图 1-8 所示。$A_{260}$ 急剧变化的温度范围一般为 6～8℃。当 $A_{260}$ 增加到最大增值的一半时，即相当 $A_{260}$ 值达到约 1.185 时，这时的温度叫做 DNA 的熔解温度或熔点，用 $T_m$ 表示。这是一个鉴定 DNA 的非常有用和非常方便的参量。

常用的 DNA 变性方法主要是热变性方法和碱变性方法。热变性使用得十分广泛，特别是用于变性动力学的研究。然而高温可能引起磷酸二酯键的断裂，得到长短不一的单链 DNA。而碱变性方法则没有这个缺点，在 pH11.3 时，全部氢键都被淘汰，DNA 完全变成单链的变性 DNA。在制备单链 DNA 时，优先采取这种方法。要维持单链状态，或者保持 pH 大于 11.3，或者盐浓低于 0.01mol/L。在这样低盐浓度时，由于磷酸基的静电斥力，可以配对的碱基无法相互靠近，碱基堆集作用也仍然保持在最低水平。

## 二、复　性

已经变性的 DNA 需要下列条件才能复性：(1) 有足够的盐浓度以消除磷酸基的静电斥力，常用的盐浓度为0.15至0.50mol/L的NaCl；(2) 有足够高的温度以破坏无规则的链内氢键，但又不能太高，否则配对碱基之间的氢键又难以形成的。一般使用比 $T_m$ 低 20～25℃ 的温度。

复性机制　复性并不是两条单链重新缠绕这样的简单过程。它首先从单链分子的无规则碰撞运动开始，这种碰撞过程是随机的，与 DNA 的浓度，溶液的温度和离子强度有关。有时两条单链之间非对应的某一小段碱基之间也能形成氢键，但总的来说，这样的安排使链上的大部分碱基都不能互补。这样的碰撞可能多次发生，但所形成的氢键都是短命的，很快会被分子的热运动所瓦解。只有当应该配对的一部分碱基相互靠近时，一般认为需要 10～20 个碱基对，特别是富含 G·C 的节段首先形成氢键，产生一个(或几个)双螺旋核心(这一步叫成核作用，nucleation)；然后，两条单链的其余部分就会像拉拉链那样迅速形成双螺旋结构(这一步叫做拉拉链作用，zippering)。因此，复性过程的限制性因素乃是分子碰撞过程，大部分变性但还没有完全变性的 DNA 会迅速地复性，也就是因为这一过程中不大需要这样一个碰撞过程。完全变性 DNA 一般需要几个小时才能复性。需要注意，复性 DNA 分子不一定是原有的一对互补链，可以说，大部分复性 DNA 分子都不是原配的。

### 三、杂　交

上述复性 DNA 中，如果两条链来源不同，就叫做杂交分子。DNA 杂交技术可以用来检验 DNA 的缺失、插入，还可以用来考察不同生物种类在 DNA 分子中的共同序列和不同的序列以确定它们在进化中的关系。这样生成的 DNA 杂交分子常称为异源双链DNA (heteroduplexes)。这一技术是将复性方法和电子显微镜方法结合在一起。

DNA 和 DNA 杂交技术以及 DNA 和 RNA 杂交技术在核酸技术中占有十分重要的地位。其基本原理就是硝酸纤维素滤膜能牢固地结合单链 DNA，而不能结合双链DNA和RNA，具体操作步骤请读者参考有关文献。

### 四、Cot 曲　线

DNA 复性是一种双分子二级反应，单链消失的速度可用下面公式表示：

$$-\frac{dC}{dt}=kC^2$$

其中 C 为单链 DNA 的浓度，单位是每升的核苷酸摩尔数；t 是时间，单位为秒；k 是二级反应常数，单位是升/mole 秒，k 值取决于阳离子浓度，温度，片断大小和 DNA 分子序列的复杂性(sequence complexity of the DNA population)。上面公式可以重排为 $-\frac{dC}{C^2}=k\,dt$。当 $t=0$ 时，$C=C_0$，将上式积分：$-\left(\frac{1}{C_0}-\frac{1}{C}\right)=kt$，即 $\frac{1}{C}-\frac{1}{C_0}=kt$，可重排成 $\frac{C}{C_0}=\frac{1}{1+kC_0t}$。

在上式中，当 $t=0$ 时，$C=C_0$，表明所有的 DNA 都是单链。$C_0$ 为 DNA 的总浓度。

复性的分数 $C/C_0$ 是起始浓度和经过时间的乘积 Cot 的函数，这样的函数可以绘成下面的图，称为 Cot 曲线。当 $C/C_0=\frac{1}{2}$ 时的 Cot 值定义为 $Cot_{\frac{1}{2}}$，因此 $Cot_{\frac{1}{2}}=\frac{1}{k}$。

各个不同的 DNA 的 $Cot_{\frac{1}{2}}$ 值是不同的，不仅与 k 值相关，而且 DNA 分子序列的复杂性影响 k 值。DNA 分子序列的复杂性 X 值定义为：最长的没有重复序列的核苷酸对的数值。例如(ATATAT…)，其 X 值为 2，$(ATGC)_n$ 的 X 值为 4，没有重复序列的含有 $10^5$ 核苷酸对的 DNA 分子的 X 值为 $10^5$。

在DNA总浓度(以核苷酸为单位)相同的情况下，片断越短，片断浓度就越高，复性所需的

时间 $t_{\frac{1}{2}}$ 也越短，$Cot_{\frac{1}{2}}$ 也越小，因此，X 值与 $Cot_{\frac{1}{2}}$ 的关系是 $X = KCot_{\frac{1}{2}}$。其中，比例常数K将取决于反应条件(阳离子浓度，片断大小等)，在通常使用的标准状况下(0.18当量的阳离子，片断大小 400 个核苷酸)，$K \approx 5 \times 10^5$(升×核苷酸对/核苷酸 mole×秒)，即在此条件下，$X = 5 \times 10^5 Cot_{\frac{1}{2}}$(核苷酸对)。

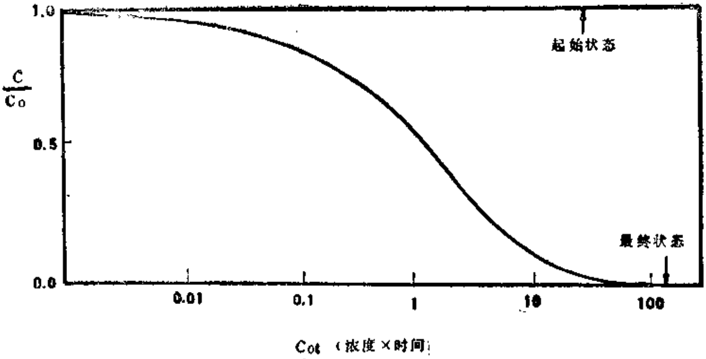



图 1-9 Cot 曲线

不同物种的核酸的 Cot 曲线示于图 1-10

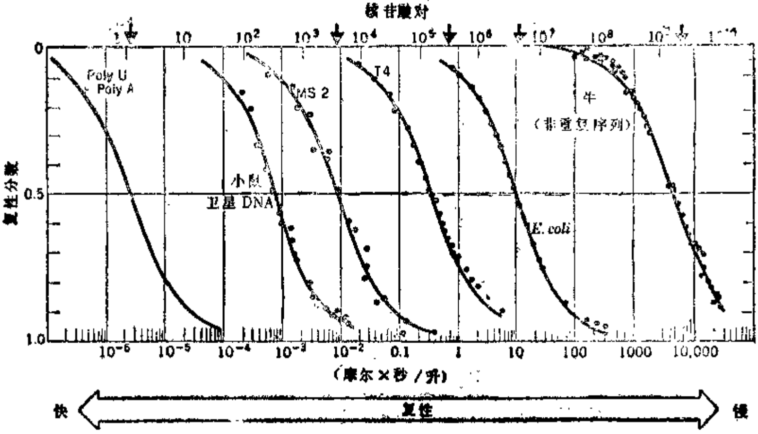



图 1-10 不同物种核酸的 Cot 曲线

图 1-10 表明，$Cot_{\frac{1}{2}}$ 是与基因组的大小成正比的。其中 polyU + polyA，其 X 值为 1 核苷酸对，因而复性最快；MS2 是 RNA 噬菌体，T4 为 DNA 噬菌体，牛的 DNA 是指它的非重复序列。每个基因组的大小用箭头示于图的上方。

## 第七节 DNA 二级结构的多形性(Polymorphism)[1,14~29]

1953 年 Watson 和 Crick 提出的 DNA 右手双螺旋模型基本上相当于 B 型。B 型是用生理盐浓度的 DNA 溶液抽成纤维，然后在 92% 的相对湿度下进行 X 光衍射测定出来的。大量的工作证明了在溶液中的 DNA 确实是采取 B 构象；细胞中含有大量的水分，据信细胞 DNA 也是主要采取 B 构象。然而，细胞并不是均一的溶液，而且细胞中的 DNA 常常与蛋白质紧密结合，在某些区域发生构象上的改变是完全有可能的。除了上述的 B 构象外，人们还发现了 A 构象、C 构象、D 构象和 E 构象等右手双螺旋构象以及左手双螺旋的 Z 构象。

A 构象是在以 $Na^+$、$K^+$ 或 $Cs^+$ 作为反离子时将 DNA 纤维在 75% 的相对湿度下进行 X-光衍射测定出来的。DNA-RNA 杂交分子，RNA-RNA 双链结构均采取 A 构象。C 构象是以 $Li^+$ 为反离子时在 66% 的相对湿度下测定出来的，双螺旋直径与 B 构象大体相同(19Å)，而每圈螺旋的碱基对数为 $9\frac{1}{3}$，目前尚无证据说明生物体内有 C 型 DNA 的存在。D 构象和 E 构象仅适用于缺少 G—C 对的 DNA 分子，每圈螺旋的碱基对数分别为 8 和 $7\frac{1}{2}$。人工组建的 poly(dA)、poly(dT)仍然是 B 构象，但其小沟明显变狭，这可能妨碍与组蛋白八聚体形成核小体结构。

1979 年，A.H.—J.Wang 等人用 X 光衍射技术对六聚体 d(CpGpCpGpCpG) 单晶的结构进行研究。出人预料地发现，这种六聚体的构象既不是B型，也不是 A 型，而是一种全新的左手双螺旋构象，叫做 Z 构象，之所以命名为 Z 型，是因为其脱氧核糖-磷酸骨架按之字形(zig-zag)路径从一端伸向另一端。

下面，我们比较一下 B-DNA，A-DNA（包括 DNA-RNA 杂交分子）和 Z-DNA 的结构特征。

B-DNA：右手双股螺旋；每圈螺旋 10个碱基对，螺旋扭角为36°，螺距34Å，每个碱基对的螺旋上升值为 3.4Å；碱基倾角 −2°，碱基平面基本上与螺旋轴垂直；螺旋轴穿过碱基对，大沟宽而略深，小沟窄而略浅。B-DNA 结构见图 1-11。

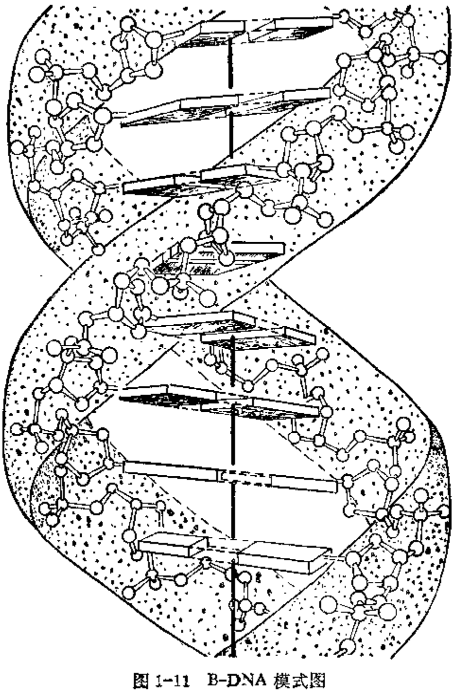


图 1-11 B-DNA 模式图

A-DNA：右手双股螺旋；每圈螺旋 10.9 个碱基对，螺旋扭角为 33°；螺距 32Å，每个碱基对的螺旋上升值为 2.9Å；碱基倾角13°，碱基平面不再与螺

旋轴垂直；螺旋轴不穿过碱基对，而是位于大沟中，碱基对在小沟中围绕着螺旋轴，形成中空结构；小沟宽而浅，大沟极深，其深度为从螺旋表面经过螺旋轴再向对面延伸一段距离。A-DNA结构见图 1-12。

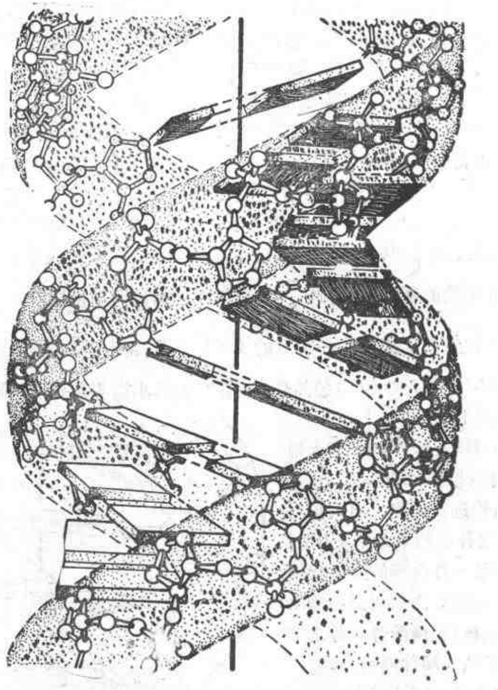


图 1-12　A-DNA 模式图

Z-DNA：左手双股螺旋；每圈螺旋12个碱基对，螺旋扭角为 $-51°$(G—C)和 $-9°$(C—G)；螺距 46Å，每个碱基对的螺旋上升值为 3.5Å(G—C) 和 4.1Å(C—G)；碱基倾角9°，即碱基平面不与螺旋轴垂直；螺旋轴不穿过碱基对，而是位于小沟中，大沟已不复存在，小沟狭而深，两条链上的磷酸基隔着狭窄的小沟面对面。Z—DNA 结构见图 1-13。

在上述三种形式 DNA 的结构特征中，沟的特征在遗传信息表达过程中起着关键的作用。调控蛋白质都是通过其分子上特定的氨基酸侧链与沟中碱基对两侧潜在的氢原子供体或受体形成氢键而识别 DNA 的遗传信息的。这些潜在的氢原子供体和受体在图 1-14 中以带星号字母表示。从图中可以看出，大沟和小沟中这些氢原子的供体和受体及其排列是不同的；在大沟中，G—C 对是 a—a—d 或 d—a—a，而 A—T 对为 a—d—a；在小沟中，G—C 对是 a—d—a，而 A—T 对为 a——a(a 表示受氢体、d 表示供氢体)。由此可见，大沟携带的信息要比小沟多。除此以外，沟的宽窄和深浅也直接影响到调控蛋白质对 DNA 信息的识别，例如在Z-DNA中，小沟狭而深，大沟消失，碱基对富于信息的一边在大沟位置暴露于螺旋表面。可见调控蛋

白质对 Z-DNA 信息的识别方式与对 B-DNA 信息的识别方式是不相同的。对于 A-DNA(包括DNA-RNA杂交双链,RNA-RNA 双链)来说,大沟极深,碱基对在小沟一边比较暴露,由此可见,调控蛋白质对 A-DNA 信息的识别方式也不同于B-DNA,二者差异可能比 B-DNA与Z-DNA 的差异还要大,尽管A-DNA 和 B-DNA 同属右手双螺旋,而 Z-DNA 为左手双螺旋。

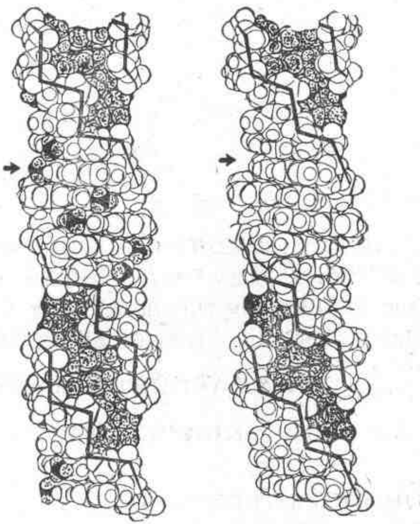


图 1-13 Z-DNA 的空间充填模型以及胞嘧啶的 5-甲基对 Z-DNA 构象的稳定作用

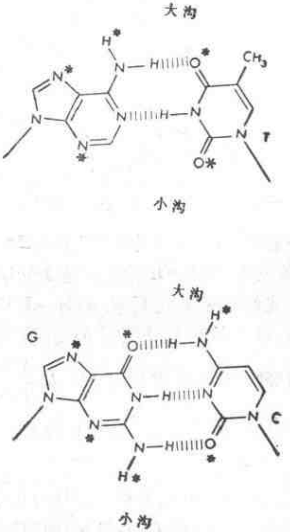



图 1-14 大沟和小沟中潜在的氢原子供体和受体(带星号字母)

沟的深与浅主要决定于螺旋轴的位置,而沟的宽窄主要决定于两个糖苷键(C—N 苷键)与碱基对的相对位置和糖苷键的顺反构象(图 1-15)。关于螺旋轴决定沟的深浅,其道理是显而易见的。螺旋轴在哪个沟中,哪个沟就深。两个糖苷键与碱基对的相对位置决定了大沟和小沟,前面已经讲过。在 B-DNA, A-DNA 等右手双螺旋构型中,糖苷键均为反式构象。而在 Z-DNA 中,嘧啶的糖苷键仍为反式构象(anti),而嘌呤的糖苷键则为顺式构象(syn)。嘧啶的糖苷键之所以不能采取顺式构象,这是因为嘧啶的第 2 位酮基不允许顺式糖苷键的存在,否则距离太近,不符合立体化学原理。这样,在 Z-DNA 中,嘧啶上的糖苷键离开小沟向外挑出,而嘌呤上的糖苷键则向小沟一边弯转。这样,由于一条链中嘧啶和嘌呤的交替排列,因而糖苷键的反式构象和顺式构象亦交替排列,从而使得 Z-DNA 的主链采取之字形的路径。从一个磷原子经过嘌呤核苷到下一个磷原子,其路经几乎平行于螺旋轴,而另一方面,从一个磷原子经过嘧啶核苷到下一个磷原子,其路径几乎与螺旋圆柱体的侧面相切,并与螺旋轴垂直。因此,虽然 Z-DNA 左手双螺旋一周有12个碱基对,但形式上只有 6 个重复单位,每个重复单位是两个碱基对。

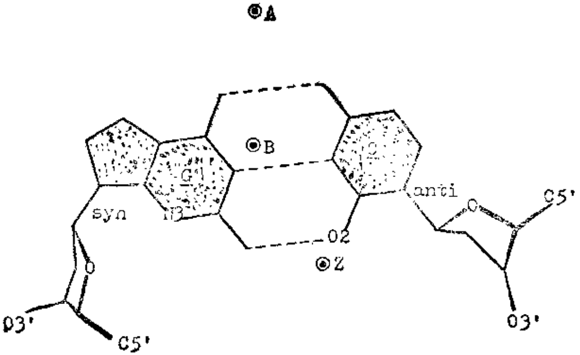



图 1-15　三种 DNA 螺旋轴的位置及 Z-DNA 的糖苷键构象

在各种形式的 DNA 中，除了 B-DNA 以外，最引人注目的就要算 Z-DNA 了。早在1972年，Pohl 等就发现，当 NaCl 浓度超过 2mol/L 或 $MgCl_2$ 浓度超过 0.7mol/L 时，poly(dG—dC) 溶液的旋光性改变符号。后来，A.H.-J.Wang 等对 d(CpGpCpGpCpG)的单晶体进行 X-光衍射分析，得出 Z 构象。后来人们发现，不但 poly(dG-dC)具有 Z 构象，在双螺旋多核苷酸链中，只要嘌呤和嘧啶相间排列，如 $\frac{\text{ACACACAC}}{\text{TGTGTGTG}}$ 等，甚至是 AT 的交替排列，在一定的条件下，都可能出现 Z 构象。在很长的 DNA 分子中，如果有一段嘌呤-嘧啶交替排列的顺序，例如：

……T G A T C | C G C G C G C G | A G T C T T ……
……A C T A G | G C G C G C G C | T C A G A A ……

在2mol/L NaCl中，打框部分可以采取 Z 构象，其余部分仍为B 构象。这一转变过程是：鸟嘌呤绕着糖苷键由反式变为顺式，而胞嘧啶连同核糖一起转了个身。正因为如此，才使主链成为之字形。这个转变机制称为翻板假说(flipping over)，见图 1-16。这种转变过程并非是单一的由

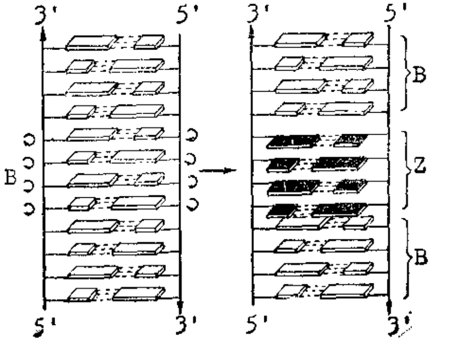



图 1-16　翻板假说示意图

一个方向转到另一方向，而是具有一系列复杂的中间过程。在生理状态下，B-DNA，A-DNA，Z-DNA 可能处在一个动态平衡状态之下。Pohl 等人测量不同盐浓度下 poly (dG—dC) 的 CD 谱就表明了 B-DNA 与 Z-DNA 的相互转变是有中间过程(图 1-17)。

A. H.-J. Wang 在低盐溶液中测量 $\begin{matrix}CGCGCG\\GCGCGC\end{matrix}$ 的 CD 谱表明并无 Z-DNA 存在，而在结晶之后则全部为 Z-DNA。人们认为，即使在溶液状态时也会有少量难以测量出来的 Z-DNA 存在，然后以此为核心形成 Z-DNA 结晶。

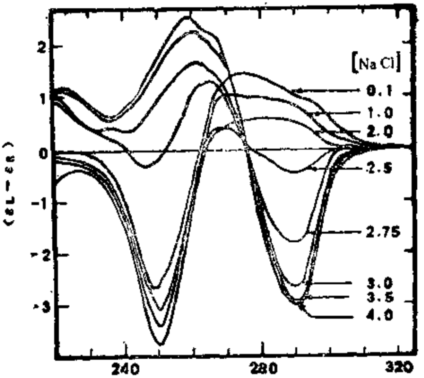


图 1-17　不同 NaCl 浓度下 poly(dG-dC) 的 CD 谱

E. M. Lafer 等人对 Z-DNA 抗体的研究表明，细胞内 DNA 分子中确实存在左手双螺旋区域。Z-DNA 具有很强的免疫原性，这与 B-DNA 明显不同，B-DNA 没有免疫原性，用通常的方法很难得到 B-DNA 的抗体。若将三分之一以上的鸟嘌呤的 C8 溴化，则能使poly(dG-dC)稳定于 Z 构象。这是因为用较大的溴原子取代了 C8 上的氢子，稳定了鸟嘌呤糖苷键的顺式构象。CD谱测量表明，这样溴化的poly(dG-dC)在低盐浓度下与没有溴化的poly(dG-dC)在4mol/L NaCl 时表现出相同的 CD 谱。将溴化的 poly(dG-dC) 注射到兔子或小鼠体内即可产生高浓度的Z-DNA抗体。这样的抗体可以在0.2mol/L NaCl 时与溴化的 poly(dG-dC)结合而沉淀；在 1.5mol/L NaCl 时，可与 $C_5$ 甲基化的 poly (dG-dC) 结合；只有在 4mol/L NaCl 时才能与未经修饰的 poly(dG-dC) 结合。而对于提纯的 *E. coli* DNA，在任何盐浓度下都不能与 Z-DNA 抗体结合。这充分说明 Z-DNA 抗体的专一性。人们用间接免疫-电子显微镜术观察果蝇的多线染色体，表明 Z-DNA 抗体主要结合于带间区(即亮带)。

仅就生理盐浓度来说，Z-DNA 是不可能稳定存在的，其主要原因是两条主链上带负电荷的磷酸基团相距太近以致发生静电排斥。然而，细胞内确实含有促进 B-DNA 向 Z-DNA 转变的因素：

(1) 胞嘧啶 $C_5$ 的甲基化。1981 年 Behe 等人发现，胞嘧啶 $C_5$ 的甲基化使poly(dG-dC)在生理盐浓度下就稳定地采取 Z 构象，其原因是甲基周围形成一个局部的憎水区域，有利于 Z-DNA的稳定；而对于B-DNA，这一甲基化的憎水区域伸到大沟中，反而不利于B-DNA 的稳定，促使 B-DNA 向 Z-DNA 的转变。而在真核生物中确实存在着 CG 序列中胞嘧啶 $C_5$ 甲基化的现象，人们还发现，甲基化程度越高，基因转录活性越低，而去甲基化则与转录激活相联系。

(2) 阳离子的结合。阳离子结合在带负电荷的磷酸基的周围，可大大降低磷酸基的静电排斥作用。$Mg^+$ 比 $Na^+$ 更为有效，而多胺化合物，如精胺，亚精胺，其效力更强。生物体内确有多胺化合物的存在。

(3) Z-DNA 结合蛋白：Z-DNA 抗体是第一个被证明有稳定 Z-DNA 作用的蛋白质，它可以在低盐浓度下促进 B-DNA 向 Z-DNA 的转变，其机制可能就是上文所述的翻板机制。

1982 年，Nordheim 等人应用 Z-DNA 亲和层析柱分离出许多能结合于 Z-DNA 但不能结合于 B-DNA 的蛋白质。这些蛋白质是不均一的，它们是否具有"翻板"能力尚未经过检定。

(4) 负超螺旋。所有生物的 DNA 都具有负超螺旋。可以用 DNA 拓扑异构酶控制DNA的负超螺旋数，然后测量其与 Z-DNA 抗体的结合情况。利用质粒将 poly(dG-dC) 插入其中，只有存在负超螺旋时，才能与 Z-DNA 抗体结合。质粒 pBR322 具有一段 14 个碱基对的嘌呤-嘧啶交替排列区(只有一个碱基对例外)；通过抗体结合试验，表明它确实采取Z构象。这是第一个探明顺序的天然 Z-DNA 片段。其序列是 1448CACGG*GTGCGCATG1461，其中G*为不符合嘧啶-嘌呤交替排列规律的一个碱基对。人们还发现啮齿动物的微小病毒(parvo-virus)的 DNA 含有两段 C·G 交替排列的序列，它们与复制起点相邻。

上述作用方式中，有些很象大分子蛋白质的别构效应 (allosteric effect)。DNA 大分子是否也存在着别构效应，其效应物 (effector) 是什么以及如何调节，都有待进一步研究。

除了 Z-DNA 以外，是否还存在着其他类型的左手双螺旋 DNA 呢？通过 Brahms 等人对 V 形 DNA 的研究，看来似乎只有 Z-DNA 一种左手双螺旋构象。

附带说一下，为什么 RNA-RNA 双链和 RNA-DNA 杂交分子都只能采取 A 构象而不能采取 B 构象呢？这是因为 RNA 的核糖 2′-OH 的存在，使之不能适应 B 构象，否则它将与相邻的磷酸，核糖和碱基上的若干原子靠得太近而造成不安定状态(图 1-18)。

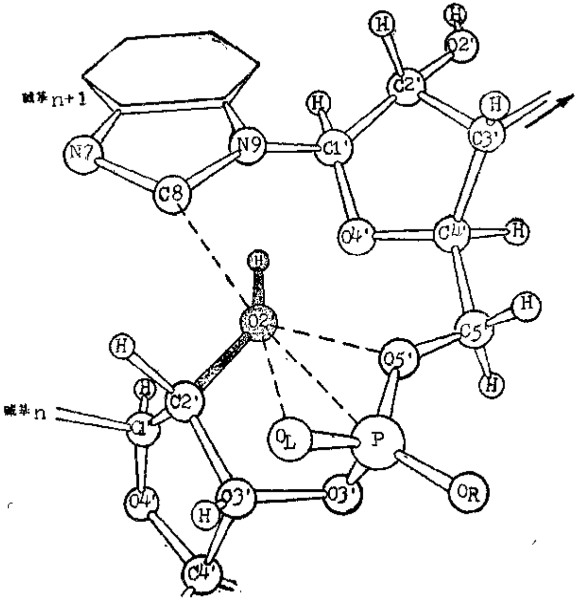



图 1-18　RNA 核糖的 2′-OH 在 B 构象中造成的不安定状态

# 第八节 DNA 构象家族[1,6,30,31]

上述三种主要类型的 DNA,即 A-DNA,B-DNA,Z-DNA,都不是只代表单一的构象,而是代表了一组相关的构象。这样一组相关的构象可以称为构象家族(comformational family)。随着 DNA 分析技术由 DNA 纤维的 X 光衍射发展到单晶 DNA 的 X 光衍射,人们能够探知 DNA 构象的局部变化。用 DNA 纤维进行 X 光衍射得出的数据只能代表一组构象的平均值,用 DNA 单晶进行 X 光衍射,得出的构象家族的平均值也差不多,这并不使人惊讶。使人惊讶的是这些数值的变化范围,例如每个碱基对的螺旋扭角对于 B-DNA 为28～42°,对于 A-DNA 为 16～44°,为什么会发生这样的变化范围呢?其根源在于一个碱基对中两个碱基的螺旋桨扭转效应(propeller twist)。在对 DNA 纤维进行 X 光衍射的研究中,基本上忽视了螺旋桨效应。

所谓螺旋桨扭转就是一个碱基对中的两个碱基并不处于同一平面中,而是两个碱基平面相对碱基的长轴各自向着相反的方向扭转。如果沿着碱基对的长轴看去,靠近的一个碱基总是顺时针方向扭转。这样,螺旋桨扭角也总是定义为正值(图 1-19),A-DNA 为15°,B-DNA 为 12°,个别情况下可以低至 3°,高至 25°。

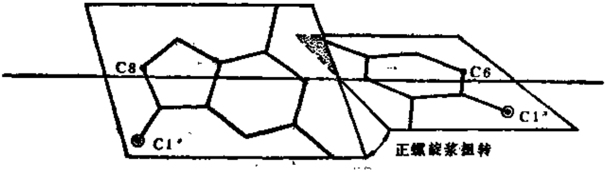



图 1-19 碱基对的正螺旋桨扭转效应

螺旋桨扭转效应的后果是:一方面改善了同一条主链中相邻碱基之间的堆集作用,使相邻碱基之间的重叠面积增大;另一方面又导致了相邻碱基对中但位于两条主链上的嘌呤环的过份接近,产生了构象上的不安定状态。原因是:嘌呤环是双环,它的长度接近或超过螺旋轴心,而每个碱基对的螺旋桨扭转又均为正值,这必然引起两条链上位于相邻碱基对中的嘌呤环之间的抵牾。随着 DNA 的碱基序列不同,嘌呤环的抵牾状况也不同。当相邻二核苷酸对为嘌呤-嘧啶顺序时,抵牾发生在大沟一侧;当相邻二核苷酸对为嘧啶-嘌呤顺序时,抵牾发生在小沟一侧,且小沟抵牾状况更为严重(图 1-20)。据 Calladine 计算,小沟抵牾效应大约是大沟抵牾效应的两倍。

解救这种构型上的危机,最简单的方式是压缩相邻二核苷酸对中一个碱基对或两个碱基对的螺旋桨扭转,但从实际的空间构象上来说,采取其他两种方式更为容易些。第一种方式是增加抵牾部位的碱基对平面转角(简称碱基转角,base roll angle)。尽管碱基对的两个碱基平面之间有螺旋桨扭转,但人们还是把碱基对看成一个整体,通过碱基对的长轴取两个碱基平面之间的平均平面作为碱基对的平面。两个相邻碱基对平面之间的夹角就称为碱基转角。如果这一夹角张开于小沟方向,碱基转角定义为正值;如果这一夹角张开于大沟方向,则碱基转角定义为负值。当发生大沟抵牾时,则碱基转角向着大沟张开,其负值增加;当发生小沟抵

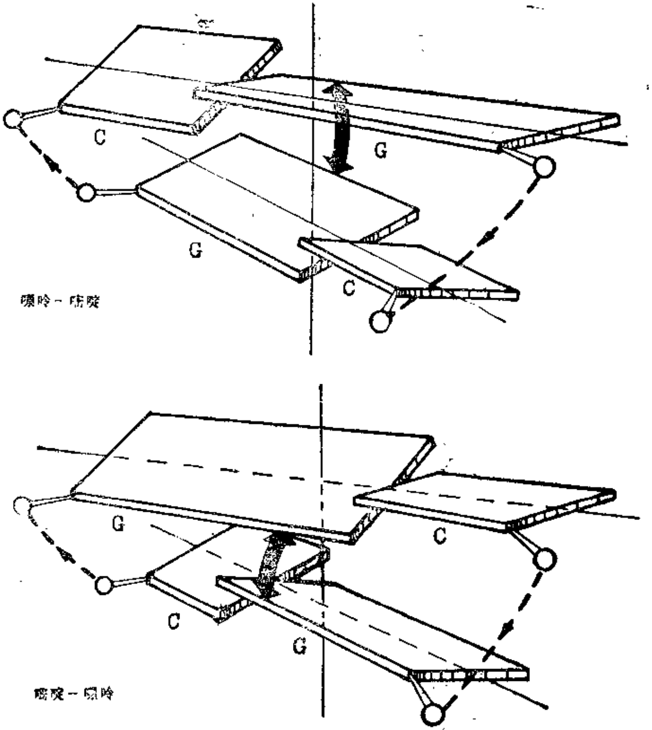



图 1-20　不同碱基序列的嘌呤环抵牾状况

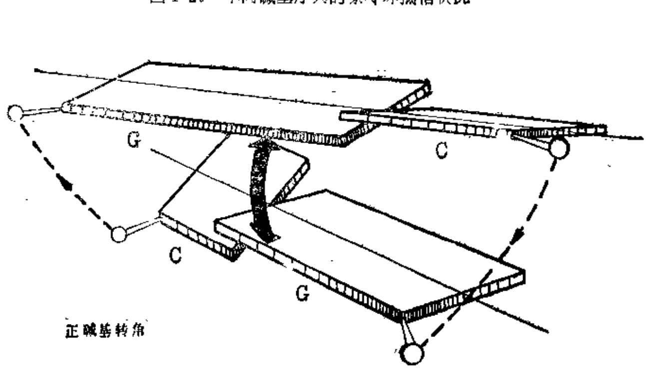



图 1-21　增加碱基转角以解救嘌呤环抵牾危机（本图所示为小沟抵牾的解救）

牾时，则碱基转角向着小沟张开，其正值增加，而且其增加的数值大约是大沟抵牾时负值增加的两倍(图 1-21)。另一种解救方式是减少螺旋扭角 (helical twist angle)。不管是大沟抵牾还是小沟抵牾，其结果都是降低螺旋扭角，只是小沟抵牾引起螺旋扭角降低的数值是大沟抵牾的两倍。根据 Calladine 的计算方法，人们可以比较精确地预测两种右手双螺旋结构的碱基转角，螺旋扭角以及螺旋桨扭角本身这些重要的螺旋参数。这些参数的预测值和实际测量值相当吻合。现将三种主要 DNA 构象的螺旋参数及其变化范围列于表 1-2。

表 1-2 三种主要 DNA 构型的螺旋参数(括号内的两组数值为实际观察到的变化范围)

| | A-DNA | B-DNA | Z-DNA |
|---|---|---|---|
| 手　　性 | 右 | 右 | 左 |
| 螺旋扭角($\phi°$)<br>(helical twist angle) | 33.1±5.9<br>(16.1−44.1) | 35.9±4.3<br>(27.7−42.0) | −51.3±1.6(G-C)<br>−8.5±1.1(C-G) |
| 每圈螺旋的碱基对数目 | 10.9 | 10.0 | 12.0 |
| 每个碱基对的螺旋高(Å) | 2.92±0.39 | 3.36±0.42 | 3.52±0.22(G-C)<br>4.13±0.18(C-G) |
| 碱基倾角(°)<br>(base inclination angle) | 13.0±1.9 | −2.0±4.6 | 8.8±0.7 |
| 螺旋桨扭角(°)<br>(propeller twist angle) | 15.4±6.2 | 11.7±4.8 | 4.4±2.8 |
| 碱基转角(°)<br>(base roll angle) | 5.9±4.7 | −1.0±5.5 | 3.4±2.1 |

## 第九节　DNA 超螺旋和拓扑异构现象[10,32~41]

自从 1965 年 Vinograd 等人发现多瘤病毒的环形 DNA 的超螺旋以来，现已知道绝大多数原核生物都是共价封闭环(covalently closed circle)简称 CCC 分子。这种双螺旋环状分子再度螺旋化成为超螺旋结构(superhelix 或 supercoil)，如图 1-22 所示。有些单链环形染色体(如 $\phi$×174)或双链线形染色体(如噬菌体 λ)，在其生活周期的某一阶段，也必须将其染色体变为超螺旋形式。对于真核生物来说，虽然其染色体多为线形分子，但其 DNA 均与蛋白质相结合，两个结合点之间的 DNA 形成一个噜噗(loop)结构，类似于 CCC 分子，同样具有超螺旋形式。

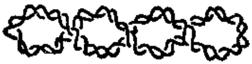


图 1-22　DNA 分子的超螺旋

超螺旋是有方向的，有正超螺旋和负超螺旋两种。为了说明超螺旋的方向，我们可以用一根细绳做下述实验。原来的绳子的两股以右旋方向缠绕。如果在一端使绳子向紧缠方向(overwinding)捻转，再将绳子两端连接起来，则会产生一个左旋的超螺旋以解除外加的捻转造成的胁变。这样的超螺旋叫做正超螺旋。相反如果在绳子一端向松缠方向(underwinding)捻转，再将绳子两端连接起来，则会产生一个右旋的超螺旋以解除外加的捻转所造成的胁变。

这样的超螺旋叫做负超螺旋。对于右手螺旋 DNA 分子来说，如果每圈初级螺旋的碱基对数小于 10.5，则其二级结构处于紧缠状态，由此而产生的超螺旋就是正超螺旋；如果每圈初级螺旋的碱基对数大于 10.5，则其二级结构处于松缠状态，由此而产生的超螺旋就是负超螺旋。由此可见，超螺旋总是要向着抵消初级螺旋改变的方向发展。

可以用以下的术语描写 DNA 的螺旋状态：

(1) Vinograd 方程：Vinograd 等人对封闭的双链环状 DNA 分子的螺旋性质进行研究，认为 DNA 超螺旋是与它的初级螺旋密切相关的。这个关系可以用如下公式表示：$\alpha=\beta+\tau$，其中 $\alpha$ 称为拓扑缠绕数(topological winding number)，$\beta$ 为初级螺旋的圈数，$\tau$ 为超螺旋圈数。

(2) 带子模型 (ribbon model)：White 差不多在同时(1969 年)对于封闭的带子(closed ribbon)进行微分几何学的研究，得到如下算式 $L_k=Tw+Wr$。式中 $L_k$ 称为链环数(linking number)，即带子两个边缘相交的次数，对于 DNA 分子来说，即为两条链相交的总次数。习惯上，将右手双螺旋 DNA 的链环数定义为正值。$Tw$ 称为缠绕数(twisting number)，对于 DNA 分子来说即为一条链绕着另一条链所形成的初级螺旋的圈数。$Wr$ 为螺旋轴的一种性质，后来(1971 年)Fuller 称之为扭曲数(writhing number)，它代表了双螺旋分子在空间上相对于双螺旋轴心的扭曲，在直观上等于超螺旋数目。带子模型提供了一个分析超螺旋 DNA 性质的方法，其数学概念也更为严密。一个封闭的环状双链 DNA 分子，当其主链的共价键不发生断裂时，它的 $L_k$ 值是不会改变的，但其 $Tw$ 值和 $Wr$ 却是可以改变的。同样，当一个 DNA 分子受到一个外加的断裂-再连接作用如拓扑异构酶的作用而 $L_k$ 值发生改变时，这一改变可以归结为 $Tw$ 值的改变或 $Wr$ 值的改变或二者共同改变的结果，写成公式就是 $\Delta L_k=\Delta Tw+\Delta Wr$。

由于 $\alpha=\beta+\tau$ 和 $L_k=Tw+Wr$ 两个方程式形式相同，以致于许多人误认为 $\beta$ 与 $Tw$ 相同，$\tau$ 与 $Wr$ 相同。实际上，在上述两个方程式中，$\alpha$ 和 $L_k$ 的确是相同的，是可以互换的；但是 $\beta$ 不一定等于 $Tw$，$\tau$ 也不一定等于 $Wr$。因为 $Tw$ 和 $Wr$ 均为变量，而 $\beta$ 实际上为常量 $\alpha^0$，即封闭的环形 DNA 分子处于松弛状态(relaxed form)时的链环数。实际上，Vinograd 方程是把超螺旋定义为链环数之差，即 $\tau=\alpha-\alpha^0$，这一公式的重要性在于使超螺旋数可以通过实验进行测定。尽管 $Tw$ 和 $Wr$ 在数学定义上严格，在理解 DNA 螺旋性质极为有用，但迄今为止人们还是无法从实验上直接测定 $Tw$ 和 $Wr$ 的数值，由于使 $\alpha^0$ 值增加 1 的碱基对数 $h^0$ 可以用多种方法测定出来，因此任何一个 DNA 分子的 $\alpha^0$ 都可以计算出来即 $\alpha^0=n/h^0$，(n 为总的碱基对数)。对于典型的核苷酸序列，在生理条件下，$h^0$ 值大约为 10.5 碱基对，这里条件是很重要的。上文已经说过，只要 DNA 分子的主链的共价键没有断裂，$L_k(\alpha)$ 值就不会改变。也就是说，$L_k$ 是拓扑学上不变的数值。然而 $\alpha^0$ 并不是拓扑学上不变数值，因为 DNA 分子周围条件变化，$h^0$ 则随着变化，$\alpha^0$ 也必然随着变化。也就是说，DNA 分子在一种条件下是松弛形式，而在其他不同的条件下就不一定是松弛形式；不需要主链的断裂，DNA 分子就可以采取不同的 $\alpha^0$ 值。

Vinograd 还引入了超螺旋密度的概念，即 $\sigma=\tau/\alpha^0=(\alpha-\alpha^0)/\alpha^0$，即每圈初级螺旋的超螺旋数，这一数值又称为链环数比差(specific linking difference)。对于大多数生物而言，这一数值都大约为 −0.05，这就是说，每圈右手初级螺旋结构就有 0.05 个负超螺旋。

所有的超螺旋，不管是负超螺旋还是正超螺旋，都比松弛形式含有更多的自由能。由此可

以想象，引入超螺旋的过程都是耗能的过程。下面讲拓扑异构酶时还要讲到。根据大小相同而仅仅是链环数不同的 DNA 拓扑异构体的玻耳茨曼分布就可以计算出自由能 $\Delta G_\tau$。$\Delta G_\tau$服从于如下的公式：$\Delta G_\tau=(K/N)\tau^2$ 或 $\Delta G_\tau=NK'\sigma^2$。其中 N 为 DNA 环形分子的碱基对数，而 K 和 K′ 为与 DNA 大小(当大于数千个碱基对时)无关的常数，据计算，$K/RT=1100$，$K'/RT=10$。

由上面我们可以看到，超螺旋概念的基本要点是：

(1) 初级螺旋处于松缠或紧缠状态；

(2) 与松弛形式相比具有额外的自由能。这两点实际上是一种物理状态的两种表现。所有的生物的 DNA 都具有负超螺旋，它与许多生命过程密切相关。

在溶液中和细胞内负超螺旋会部分地转变为单链泡状结构(见图 1-23)，这种单链泡状结构也是解除松缠作用造成的胁变的一种途径。要解除松缠作用造成的胁变，改变每圈初级螺旋的碱基对的数目从热力学上来说是不可能的；造成一个或几个大的单链区域从热力学上来说也是不利的；最可能的是以负超螺旋为主要形式，一小部分为单链泡状结构，二者处于平衡状态。富含 A·T 的节段或温度升高，都有利于单链泡状结构的形成。因此，由于双链 DNA 的"呼吸作用"，其负超螺旋的链环数也会有所波动。这样，在任何时刻，负超螺旋 DNA 中单链部分的比重比其他任何双链形式 DNA 都要大。在有蛋白质分子与单链 DNA 结合时，则单链泡状结构较为稳定。这种单链泡状结构，对于 DNA 复制，基因的转录以及 DNA 重组等过程的起始起重要作用。这也许就是为什么在生物内 DNA 总是采取负超螺旋形式的主要原因。

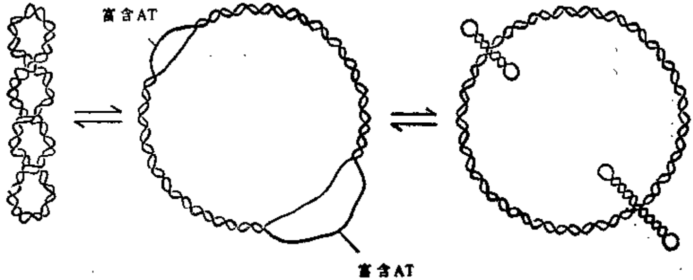



图 1-23 负超螺旋和单链泡状结构的转变。如果在单链泡状区域存在着反向重复序列，还有可能形成较为稳定的十字结构。

在真核生物中，DNA 与组蛋白质八聚体形成核小体结构，同样也存在着负超螺旋。但这种负超螺旋与在溶液中的负超螺旋的存在形式不同，手性亦不相同。这将在下一章讲述。

DNA 超螺旋是如何产生的呢？所有的 DNA 超螺旋都是由 DNA 拓扑异构酶产生的。在原核生物中，负超螺旋主要由 DNA 旋转酶(DNA gyrase)引入到已存在的双链封闭环状分子中，这一过程需要水解 ATP 以提供能量。在缺乏 ATP 时，DNA 旋转酶也可使负超螺旋松驰，但是这个作用与超螺旋化作用比较起来要低 10 倍以上。原核细胞中，这一过程是由拓扑异构酶 I 和其他的拓扑异构酶 II 来负责的，在真核细胞中，DNA 的负超螺旋则主要是染色质的结构造成的，因为 DNA 在组蛋白八聚体外面缠绕的方向有利于 DNA 向松缠方向转

变。在核小体组装过程中，也肯定有 DNA 拓扑异构酶参加。在许多真核生物中都发现了拓扑异构酶Ⅰ和拓扑异构酶Ⅱ。

DNA 拓扑异构酶可以分两类：一类叫拓扑异构酶Ⅰ，一类叫拓扑异构酶Ⅱ。拓扑异构酶Ⅰ催化 DNA 链的断裂和重新连接，每次只作用于一条链，即催化瞬时的单链的断裂和连接，它们不需要能量辅因子如 ATP 或 NAD。*E.coli* DNA 拓扑异构酶Ⅰ又称 $\omega$ 蛋白，大白鼠肝 DNA 拓扑异构酶Ⅰ又称切刻-封闭酶 (nicking-closing enzyme)。拓扑异构酶Ⅱ能同时断裂并连接双股 DNA 链，它们通常需要能量辅因子 ATP。在拓扑异构酶Ⅱ中又可以分为两个亚类：一个亚类是 DNA 旋转酶(DNA gyrase)，其主要功能为引入负超螺旋，在 DNA 复制中起十分重要的作用。迄今为止，只有在原核生物中才发现 DNA 旋转酶，另一个亚类是转变超螺旋 DNA（包括正超螺旋和负超螺旋）成为没有超螺旋的松弛形式(relaxed form)。这一反应虽然是热力学上有利的方向，但不知道为什么它们仍然像 DNA 旋转酶一样需要ATP，这可能与恢复酶的构象有关。这一类酶在原核生物和真核生物中都有发现。现将 *E.coli* 的拓扑异构酶列于表 1-3。

**表 1-3　*E.coli* 的拓扑异构酶**

| 名　称 | 结　构 | 基　因 | 亚基分子量 (KDa) | 类型 | 松弛正超螺旋 | 萘啶酮酸抑制 | 新生霉素抑制 |
|---|---|---|---|---|---|---|---|
| Ⅰ($\omega$) | $\alpha_1$ | topA | 101 | 1 | − | − | − |
| Ⅱ(gyrase) | $\alpha_2\beta_2$ | gyrA gyrB | 105 | 2 | − | + | + |
| | | | 95 | | | | |
| Ⅱ′ | $\alpha_2\beta_2$ | gyrA gyrB? | 105 | 2 | + | + | − |
| | | | 50 | | | | |
| Ⅲ | $\alpha_1$ | ? | 75 | 1 | ? | − | − |
| 拆分酶 | $\alpha_2$ | tnpR | 21 | 1 | ? | − | − |

注：(1) 结构一栏中 $\alpha$、$\beta$ 仅代表酶的亚基组分数，并不代表其基因。

(2) 含有 Tn3，$\gamma^{\delta}$ 等转座子的 *E.coli* 菌株才有拆分酶 (resolvase)。

(3) 此表不包括 *E.coli* 噬菌体编码的拓扑异构酶。

DNA 拓扑异构酶能催化的反应很多，这里只能作简单叙述。DNA 拓扑异构酶Ⅰ对单链 DNA 的亲和力要比双链高得多，这正是它识别负超螺旋 DNA 的分子基础，因为负超螺旋 DNA 常常会有一定程度的单链区。负超螺旋越高，DNA 拓扑异构酶Ⅰ作用越快。现已知道，生物体内负超螺旋稳定在 5% 左右，低了不行，高了也不行。生物体通过拓扑异构酶Ⅰ和Ⅱ的相反作用而使负超螺旋达到一个稳定状态。现已发现，编码 *E.coli* 拓扑异构酶Ⅰ的基因 *topA* 发生突变，则会引起旋转酶基因的代偿性突变；否则，负超螺旋增高，细胞生活能力降低。拓扑异构酶Ⅰ作用的碱基序列特异性不高，但切点一定在 C 的下游方向 4 个碱基(包括 C 本身)的位置。在将 DNA 单链切断后，扑拓异构酶Ⅰ连接于切口的 5′ 端，并贮藏了水解磷酸二酯键的能量用以连接切口，因而拓扑异构酶Ⅰ的作用不需能量供应。此外，拓扑异构酶Ⅰ还能促进两个单链环的复性，其作用是解除复性过程所产生的链环数的负值压力，以使复性过程进行到底。如果在同一个单链环上一个部位切断，而使另一部位绕过切口，则可产生三叶结分子(trefoil knot)。如果有两个双链环，其中一个有一个切刻，拓扑异构酶Ⅰ则可以将切刻对面的一条链切断，使完整的双链环套进去，再连接起来而成为环连体分子(catenane)，以上这三种反应示于图 1-24。拓扑异构酶Ⅰ还能催化其他反应，这将在复制和重组的机制中再讲述。

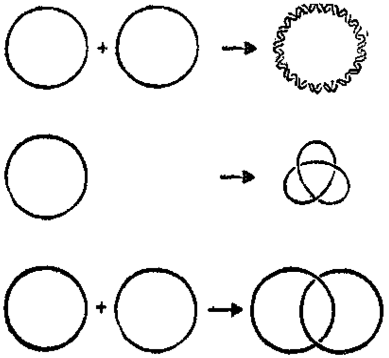



图 1-24　拓扑异构酶 I 催化的反应的一部分

大肠杆菌的拓扑异构酶 II (gyrase)除了引入负超螺旋以外,还具有形成或拆开双链DNA环连体和成结分子的能力。II类拓扑异构酶没有碱基序列特异性，它们可以和任何相交的两对双链 DNA 结合。DNA 旋转酶有两个 $\alpha$ 亚基和两个 $\beta$ 亚基。$\alpha$ 亚基约 105 KDa，为 *gyrA* 基因所编码,具有磷酸二酯酶活性，可为萘啶酮酸(nalidixic acid)所抑制。$\beta$ 亚基约 95 KDa,为 *graB* 基因所编码，具有 ATP 酶活性，可为新生霉素(novobiocin)所抑制。这两种药物均可抑制野生型大肠杆菌的 DNA 复制。可见 DNA 旋转酶为 *E.coli* 的复制所不可缺少的。在切断一条 DNA 双链后，两个 $\alpha$ 亚基各结合于切口的一个 5′ 端，并贮藏了水解磷酸二酯键而获得的能量。由于该酶的整体性，因而 DNA 链的四个断头并无任意旋转的可能性。由于酶的别构效应,使完整的双链穿过切口,然后再重新形成磷酸二酯键。$\beta$ 亚基的功能在于水解ATP以使酶分子恢复原来的构象,以便进行下一轮反应。这一点可以用 ATP 的同系物$\beta$，$\gamma$-亚氨基ATP代替 ATP 而得到证实。因为这一同系物不能被 DNA 旋转酶所水解,但它确能促进第一轮拓扑异构反应,使负超螺旋增加,而妨碍以后进一步的拓扑异构反应。

DNA 拓扑异构酶催化的反应很多,这里不可能详述。其反应本质是先切断 DNA 的磷酸二酯键,改变 DNA 的链环数之后再连接之，兼具 DNA 内切酶和 DNA 连接酶的功能，然而它们并不能连接事先已经存在的断裂 DNA,也就是说,其断裂反应与连接反应是相互耦联的。拓扑异构酶(包括 I 型和 II 型)都可以用符号转化模型进行解释(图 1-25)。

除了 DNA 拓扑异构酶可以产生异构变化以外，很多能够嵌入相邻碱基之间影响碱基堆集作用的试剂,特别是片状的染料分子,也能改变 DNA 的拓扑状态，最明显的例子就是溴化乙锭(ethidium bromide)。例如以 SV40 的CCC分子与溴化乙锭的结合试验为例,当没有染料时,此 DNA 为负超螺旋,具有较高的沉降常数(21S);当加入染料分子与核苷酸之比为0.05时,沉降数降至 16S,此时DNA 为没有超螺旋的松弛形式;当染料分子和核苷酸的比值增加到 0.09 时,沉降常数又上升到大约 21S,此时 DNA 分子为正超螺旋。这种关系如图 1-26 所示。不过要注意的是，溴化乙锭并没有改变 $L_k$ 值，只不过是由溴化乙锭分子的嵌入，增加了局部 DNA 二级结构的紧缠状态。因而，随着嵌入染料分子数的增多，起初表现为负超螺旋的减少与消失，随后便是正超螺旋的增加。这与单链 DNA 结合蛋白促进负超螺旋转变为泡状结构的情况是类似的。

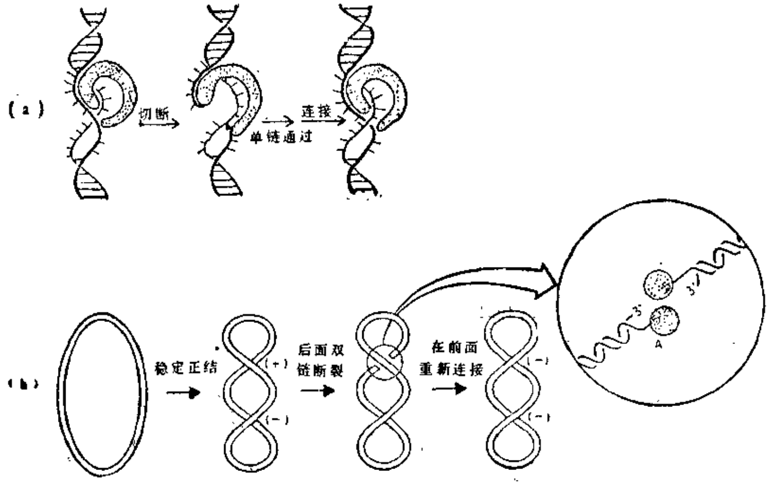



图 1-25　拓扑异构酶作用机制的符号转化模型

(a) 拓扑异构酶Ⅰ　(b) 拓扑异构酶Ⅱ

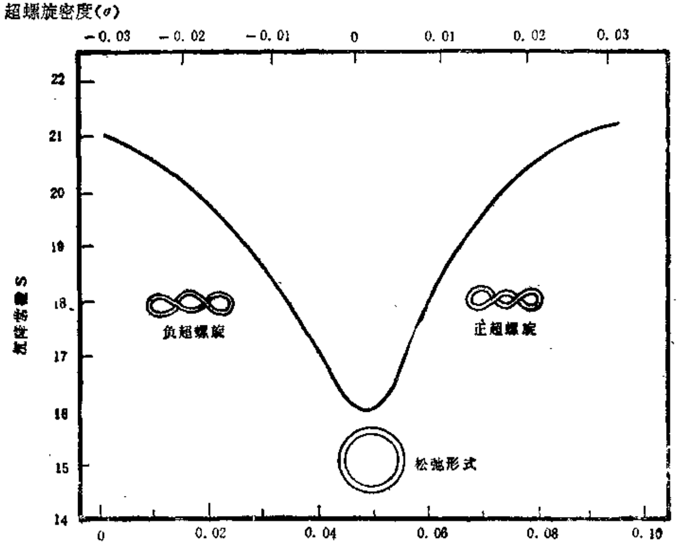



图 1-26　SV40 的拓扑结构随着溴化乙锭的结合量而改变

# 第十节　常见的 DNA 分子形式及其相互关系

现将常见的 DNA 分子形式列举如下。在谈到沉降常数时，以相同核苷酸的线形双螺旋 DNA 分子的沉降常数作为 1，其他各种分子取相对值。需要指出的是，沉降常数与电泳迁移率之间并不完全一致。

(1) I 形 DNA：具有负超螺旋或正超螺旋的双链封闭环状分子。相对沉降常数1.41。

(2) $I^0$ 形 DNA：没有超螺旋的双链封闭环状分子，也就是松弛形式 (relaxed form)，相对沉降常数 1.14。

(3) Ⅱ 形 DNA：在一条链上或两条链上有一个或几个切刻的双链环状分子，又称为切刻环(nicked circle)，相对沉降常数 1.14。

(4) Ⅲ 形 DNA：线形双螺旋分子。相对沉降常数 1.00。

(5) 坍缩 DNA (collapsed DNA)：当 I 形或 $I^0$ 形 DNA 在碱变性或热变性时，其氢键断裂，而两条链无法分离，结果生成两条链紧密缠结的分子。坍缩 DNA 具有异常高的沉降常

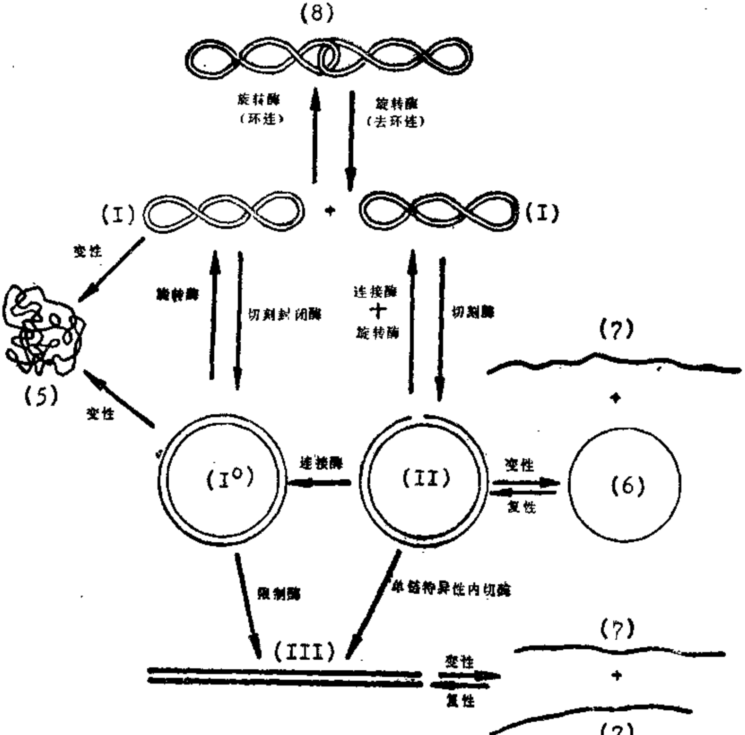



图 1-27　几种常见的 INA 分子形式的相互转变

数，相对沉降常数约3.0。

(6) 单链环状DNA：相对沉降常数大约为1.14。

(7) 线性单链DNA：相对沉降常数为1.30。

(8) 环连DNA(catenated DNA, catenane)：这是在DNA复制过程中的产物或由DNA旋转酶催化生成，由两个以上I形DNA分子环连而成。

以上八种DNA分子形式的转变表示于图1-27，其中前四种仍标以I、I°、Ⅱ、Ⅲ；后四种则分别以(5)、(6)、(7)、(8)表示。

在这些形式的分子中，I形DNA分子的特点是值得注意的：

(1) 由于超螺旋DNA分子的结构紧密，有较大的沉降常数，有较大的电泳迁移率。

(2) 与溴化乙锭等染料结合量受到分子内紧张性所限制，而Ⅱ形和Ⅲ形DNA分子没有这种限制，能够结合更多的染料分子。由于溴化乙锭的比重很小，因此在进行CsCl密度梯度平衡离心时，I形DNA的比重约为1.59，而Ⅱ形和Ⅲ形DNA的比重约为1.55。

(3) I形DNA对于温度和碱性pH有较强的抵抗力，不容易发生变性。这样，可以寻找一个临界条件，使得Ⅱ形和Ⅲ形DNA发生变性，而I形DNA不变性或很少变性。很多分离或鉴定I形DNA的方法都是依据这三点特性而设计的。

## 提　　要

DNA是几乎所有生物遗传信息的携带者。DNA不但携带着基因的编码信息，而且还贮存着基因选择性表达的指令。

DNA是由脱氧核糖核苷酸通过3′,5′磷酸二酯键连接起来的高聚物。DNA与RNA的最大区别是核糖第2位上氧原子的有无。脱氧核糖使DNA的主链的磷酸二酯键更为稳定，然而却降低了糖苷键的稳定性。这样，几乎所有生物的DNA都是以碱基互补的双链形式存在的。

对于DNA双链形式的二级结构，Watson和Crick于1953年提出了右手双螺旋模型。由脱氧核糖和磷酸基通过3′,5′磷酸二酯键连接起来的两条主链以反平行的方式和右手方向相互缠绕，构成直径为20Å的双螺旋结构。主链由于其亲水性而处于双螺旋的外表面；而碱基由于其一定程度的疏水性而处于双螺旋的内部。两条链上的碱基按A:T和G:C的互补规律相互以氢键连接。双螺旋的螺距为34Å，其中包含10个核苷酸对。两条主链并不充满双螺旋的空间，而在表面形成大沟和小沟。大沟是调控蛋白质识别DNA信息的主要场所。维系DNA二级结构的主要作用力是氢键和碱基堆集力；而磷酸基的负电荷和碱基内能则不利于双螺旋结构的稳定，在生理状况下，双螺旋的碱基对之间氢键不断地发生断裂和再生，这就是DNA的所谓呼吸作用。DNA在热或其他变性剂作用之下，双链发生分离，即变性作用。变性的DNA单链在适合的条件下又能恢复双螺旋结构，即复性作用。基于DNA的变性作用和复性作用，产生了十分有用的分子杂交技术。

DNA链上的四种碱基远非均匀分布，因而产生了一些特异的序列。其中最重要的是回文序列，许多限制性内切核酸酶和调控蛋白质的识别位点都是回文序列。富含A/T序列也是许多分子遗传学过程所不可缺少的。甚至嘌呤和嘧啶的排列顺序对双螺旋的稳定性都有重大影响。正是这种非均一性，使得每种DNA构象代表着一个构象家族，其具体参数随碱基序列而异。产生这种差异的根源在于碱基对中两个碱基平面之间的螺旋桨扭转效应。在生理状态

下可能主要为 B 构象以及少量的 A 构象和 Z 构象。Z 构象是唯一存在的左手双螺旋构象。

当右手双螺旋 DNA 的二级结构处于松缠状态时，则会形成负超螺旋。生物体内 DNA 的拓扑状态的改变是由拓扑异构酶进行的。拓扑异构酶 I 断裂和连接一条 DNA 链，每次作用改变一个链环数；拓扑异构酶 II 能同时断裂和连接双股 DNA 链，每次作用改变两个链环数。几乎所有生物的 DNA 都含有大约 5% 的负超螺旋；维持这样水平的负超螺旋是 DNA 复制，基因转录等生命活动所必需的。超螺旋 DNA 对于变性剂有较高的抗性并具有较高的沉降常数和电泳迁移率。这常常是实验室中分离纯化并鉴定超螺旋 DNA 的重要依据。

## 参 考 文 献


1. Dickerson, R. E. (1983) "The DNA Helix and How It Is Read." *Sci. Amer.* 249: 87～102.
2. Dandanell, G. et al. (1987) "Long-range Cooperativity between Gene Regulatory Sequences in a Prokaryote." *Nature* 325: 823～826.
3. Watson, J. D., and Crick, F. H. C. (1953) "Molecular Structure of Nucleic Acids, a Structure for Deoxyribose Nucleic Acid." *Nature* 171: 737～738.
4. Wilkins, M. H. F et al. (1953) "Molecular Structure of Deoxypentose Nucleic Acid." *Nature* 171: 738～740.
5. Watson, J. D., and Crick, F. H. C. (1953) "Genetic Implications of the Structure of Deoxyribonucleic Acid." *Nature* 171: 964～967.
6. Saenger, W. (1984) *Principles of Nucleic Acid Structure*, Springer-Verlag New York Inc.
7. Lilley, D. M. J. (1980) "The Inverted Repeat as a Recognizable Structure Feature in Supercoiled DNA Molecules." *Proc. Nat. Acad. Sci.* 77: 6468～6472.
8. Courey, A. J. and Wang, J. C. (1983) "Cruciform Formation in a Negatively Supercoiled DNA May be Kinetically Forbidden under Physiological Conditions." *Cell* 33: 817～829.
9. Gotoh, O., and Takashira, Y. (1981) "Stabilities of Nearest-neighbor Doublets in Double-helical DNA Determined by Fitting Calculated Melting Profiles to Observed Profiles." *Biopolymers* 20: 1033～1042.
10. Record, M. T. et al. (1981) "Double Helical DNA Conformations, Physical Properties and Interactions with Ligands." *Ann. Rev. Biochem.* 50: 997～1024.
11. Marmur, J. and Doty, P. (1962) "Determination of the Base Composition of Deoxyribonucleic Acid from its Thermal Denaturation Temperature." *J. Mol. Biol.* 5: 109～118.
12. Marmur, J. et al. (1963) "Denaturation and Renaturation of Deoxyribonucleic Acid." *Prog. Nucleic Acid Res.* 1: 231～300.
13. Spatz, H. C., and Crothers, D. M. (1969) "The Rate of DNA Unwinding." *J. Mol. Biol.* 42: 191～219.
14. Dickerson, R. E. (1983) "Base Sequence and Helix Structure Variation in B and A DNA." *J. Mol. Biol.* 136: 419～441.
15. Wang, A. H.-J. et al. (1982) "Molecular Structure of the Octamer d(GGCCGGCC-): Modified A-DNA." *Proc. Nat. Acad. Sci.* 79: 3968～3972.
16. Conner, B. N. et al. (1984) "Helix Geometry and Hydration in an A-DNA Tetramer: CCGG." *J. Mol. Biol.* 174: 663～695.
17. Alexeev, D. G. et al. (1987) "Poly (dA). Poly (dT) Is a B-type Double Helix with a Distinctively Narrow Minor Groove." *Nature* 325: 821～823.
18. Wang, A. H.-J. et al. (1979) "Molecular Structure of a Left-handed DNA Fragment at Atomic Resolution." *Nature* 282: 680～686.
19. Rich, A. et al. (1984) "The Chemistry and Biology of Left-handed Z-DNA." *Ann. Rev. Biochem.* 53: 791-846.
20. Doerfler, W. (1983) "DNA Methylation and Gene Activity." *Ann. Rev. Biochem.* 52: 93～124.
21. Pohl, F. M., and Jovin, T. M. (1972) "Salt Induced Co-operative Conformational Change of a Synthetic DNA, Equilibrium and Kinetic Studies with Poly(dG-dC)" *J. Mol. Biol.* 67: 375～396.
22. Lafer, E. M. et al. (1983) "Antibody Recognition of Z-DNA." *Cold Spring Harbor Symp. Quant. Biol.* 47: 155～162.
23. Wang, A. H.-J. et al. (1983) "Right-handed and Left-handed DNA: Structure Studies." *Cold Spring Harbor Symp. Quant. Biol.* 47: 33～44.
24. Nordheim, A. et al. (1983) "Supercoiling and Left-handed Z-DNA." *Cold Spring Harbor Symp. Quant. Biol.*

47: 93~100.

25. Rich, A. (1983) "Right-handed and Left-handed DNA: Conformational Information in Genetic Material" *Cold Spring Harbor Symp. Quant. Biol.* 47, 1~12.
26. Sullivan, K. M. et al. (1986) "A Dominant Influence of Flanking Sequences on a Local Structural Transition in DNA." *Cell* 47, 817~827.
27. Griffith, J. et al. (1986) "Visualization of the Bent Helix in Kinetoplast DNA by Electron Microscopy." *Cell*. 46: 717~724.
28. Ryder, K. et al. (1986) "An Altered DNA Conformation in Origin Region I Is a Determinant for the Binding of SV40 Large T Antigen." *Cell* 44: 719~725.
29. Burkhoff, A.M., and Tullius, T.D. (1987) "The Unusual Conformation Adopted by the Adenine Tracts in Kinetoplast DNA." *Cell* 48, 935~943.
30. Levitt, M. (1978) "How Many Base-pairs Per Turn Does DNA Have in Solution and in Chromatin? Some Theoretical Calculations." *Proc. Nat. Acad. Sci.* 75: 640~644.
31. Calladine, C. R. (1982) "Mechanism of Sequence-dependent Stacking of Bases in B DNA." *J. Mol. Biol.* 161: 343~352.
32. Gellert, M. (1981) "DNA Topoisomerases." *Ann. Rev. Biochem.* 50: 879~910.
33. Gellert, M. (1981) "DNA Gyrase and Other Type II Topoisomerases." in *The Enzymes* (edited by Paul D. Boyer, Acad. Press), V.14, Part A, pp.345~366.
34. Wang, J.C. (1981) "Type I DNA Topoisomerase." in *The Enzymes* (edited by Paul D. Boyer, Acad. Press), V.14, Part A, pp.331~344.
35. Wang, J. C. (1982) "DNA Topoisomerases." *Sci Amer* 247, 94~109.
36. Wang, J.C. et al. (1983) "DNA Supercoiling and Its Effects on DNA Structure and Function." *Cold Spring Harbor Symp. Quant. Biol.* 47: 85~91.
37. Liu, L. (1983) "DNA Topoisomerases-Enzymes that Catalyze the Breaking and Rejoining of DNA." *CRC Crit. Rev. Biochem* 15: 1~9.
38. Kirchhausen, T. et al. (1985) "DNA Gyrase and Its Complexes with DNA: Direct Observation by Electron Microscopy." *Cell* 41: 933~943.
39. Rudd, K.E. et al. (1987) "*his Operon of Escherichia coli* and *Salmonella Typhimurium* are Regulated by DNA Supercoiling", *Proc. Nat. Acad. Sci.* 84: 517~521.
40. Brown, P.O., and Cozzarelli, N.R. (1979) "A Sign Inversion Mechanism for Enzymatic Supercoiling of DNA." *Science* 50: 879~910.
41. Bauer, W. and Vinograd, J. (1968) "The Interaction of Closed Circular DNA with Intercalative Dyes." *J. Mol. Biol.* 33: 141~171.

# 第二章　有机体、染色体和基因

## 第一节　原核生物和真核生物

按照细胞的结构和遗传物质在细胞内的分布，可将生命有机体划分为原核生物和真核生物两大类。原核生物没有真正的细胞核，遗传物质存在于整个细胞之中，有时虽有相对集中的区域，但并无核膜围绕。真核生物的遗传物质集中在有核膜包围着的细胞核中，并与某些特殊的蛋白质相结合成为核蛋白以构成一种细密的结构，这种结构叫做染色体。原核生物的遗传物质只以裸露的核酸分子存在，虽与少量蛋白质结合，但不形成染色体结构。然而习惯上，原核生物的核酸分子也常被人们称为染色体。

某些有机体，如噬菌体和病毒，既不是原核生物，也不是真核生物。它们是一种超分子的亚细胞生命形式，它们的繁殖必须在寄主体内进行，因而其遗传机制与其寄主密切相关，如噬菌体（即细菌病毒）适应了原核生物的遗传战略，而动物病毒和植物病毒则使用真核生物的遗传法则。

原核生物和真核生物的划分在分子生物学和分子遗传学上十分重要。例如单细胞的酵母菌的蛋白质合成与单细胞的大肠杆菌迥然不同，而与极其复杂的高等动物相似。因为酵母菌是真核生物。

现将原核细胞和真核细胞的主要差别列于表2-1。

表2-1　原核细胞和真核细胞的差别

| 特　　征 | 原　核　细　胞 | 真　核　细　胞 |
|---|---|---|
| 遗传组织 | | |
| 核膜 | 无 | 有 |
| 不同染色体数目 | 1 | >1 |
| 核小体结构 | 无 | 有 |
| 核仁 | 无 | 有 |
| 遗传交换 | 质粒介导，单向 | 配子融合 |
| 细胞结构 | | |
| 内质网 | 无 | 有 |
| 高尔基体 | 无 | 有 |
| 溶酶体 | 无 | 有 |
| 线粒体 | 无 | 有 |
| 叶绿体 | 无 | 植物中有 |
| 核糖体大小 | 70s | 80s |
| 微管 | 无 | 有 |
| 肽聚糖细胞壁 | 有（枝原体，古细菌除外） | 无 |
| 细胞的某些功能 | | |
| 吞噬作用 | 无 | 有时有 |
| 胞饮作用 | 无 | 有时有 |
| 电子传递部位 | 细胞膜 | 线粒体膜 |
| 胞质流动 | 无 | 有 |

然而，自从70年代后期起，这种两界系统理论受到了极大的挑战，它可能会逐渐让位于三界系统理论。三界系统理论并不否定上述事实，而是考虑到上述事实并结合了分子生物学上的新发现，对生命进化系统提出了新的见解。这将在本章最后一节进行阐述。

## 第二节　基因组大小与C值矛盾[2,3]

一个物种的单倍体的染色体的数目称为该物种的基因组(genome)。一个单倍体基因组的DNA含量总是恒定的，它通常称为该物种DNA的C值。不同物种的C值差异极大，最小的枝原体(mycoplasma)只有$10^4$bp，而最大的如某些显花植物和两栖动物可达$10^{11}$bp，图2-1为不同生物种类的C值范围分布图。由图可见，随着生物的进化，生物体的结构与功能越复杂，其C值就越大，例如真菌和高等植物同属真核生物，而后者的C值大得多。这一点是不难理解的，因为结构和功能越复杂，需要的基因产物种类越多，也就是说需要的基因越多，因而C值越大。

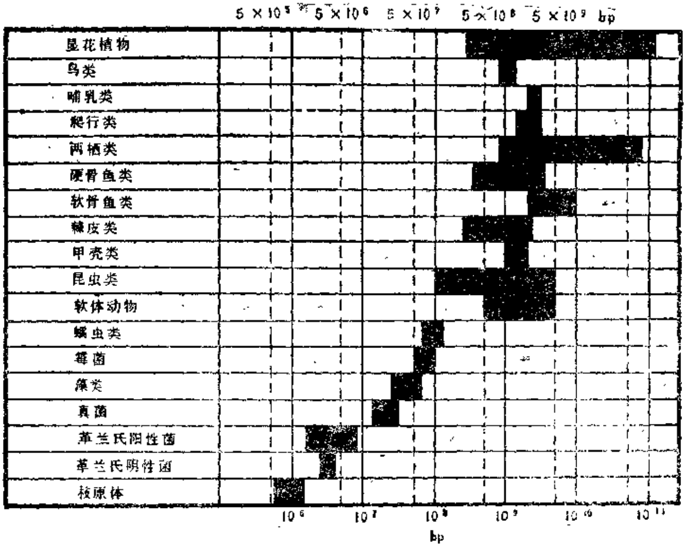



图2-1　不同生物种类基因组DNA的C值分布图

然而另一方面，在结构、功能很相似的同一类生物中，甚至在亲缘关系十分接近的物种之间，它们的C值可以相差数10倍乃至上百倍。突出的例子是两栖动物。C值小的可以低至$10^9$bp以下，C值大的可以高至几乎$10^{11}$bp。同时，哺乳动物(包括人类)的C值均为$10^9$bp的数量级，人们很难相信两栖动物的结构和功能会比哺乳动物更复杂。就哺乳动物而言，由于哺乳动物的基因含有intron即所谓"内含子"，因而基因可长达5 000～8 000 bp，少数可达10 000 bp。即使按这样大小的基因进行推算，哺乳动物的基因组相当于400 000～600 000个基因。这是可能的吗？现在虽然对哺乳动物(以及所有的高等真核生物)的确切的基因数无从

知晓，但看来是不可能有这样多的基因。估计不会超过这个数字的10%。稍微直接一点的证据是，通过DNA与mRNA杂交试验，在特定类型的细胞中表达的基因数目大约是10 000个，但各种细胞表达的基因并不相同，估计要乘上一个系数(3~4)才能得到基因组的基因数目，因而看来可能的有功能的基因数目为30 000~40 000个。一个间接的证据是通过对果蝇突变的研究，必需基因的总数大约为5 000，其平均基因大小为2 000bp，总长度相当于$10^7$bp，刚好为基因组大小的10%。这里当然有这样的可能性，有些基因并不是果蝇活存所必须的，这些基因的突变不能引起可见的效应。然而，即使考虑这些因素，基因所占基因组的比例也不会超过20%。这样，人们无法用已知功能来解释基因组的如此之大的DNA含量，这就叫做C值矛盾(C value paradox)。剩下的关键问题就是：位于基因内的DNA成份(包括转录而不翻译的区域，如"内含子"以及编码序列的两翼序列)究竟占基因组的多大比例？这些基因中有多少是必需的基因，有多少是非必需基因？非基因的DNA成份的功能是什么？基因组DNA C值的巨大差异(如在相关的两栖动物中)对基因组的活动与功能究竟有什么影响？这些都有待于进一步的研究。

美国分子生物学家们经过两年的辩论，已决定从1988年开始对人类基因组的核苷酸序列进行为期15年的研究。这项研究对于人类了解自身，诊断和治疗遗传病，阐明细胞分化和癌症的发生，探索生命的进化规律都有着极其深远的意义。

## 第三节　原核生物染色体及其基因[2,4~8]

原核生物一般只有一个染色体即一个核酸分子(DNA或RNA)，大多数为双螺旋结构，少数以单链形式存在。这些核酸分子又大多数为环状，少数为线状。我们仅以大肠杆菌以及大肠杆菌的噬菌体$\phi\times174$和$\lambda$为代表说明原核生物染色体的概况。

### 一、大肠杆菌染色体

大肠杆菌和其他细菌一样，没有明显的核结构，但是其DNA并非散布在整个细胞内，而是有一个相对集中区域，在这一小区域中形成类核(nucleoid)。在大肠杆菌的对数生长期，每个细胞可能会有2~4个这种同样的DNA分子构成的类核。大肠杆菌的类核中，DNA成份占80%，其余为RNA和蛋白质。用RNA酶或蛋白酶处理类核，可使之由致密变得松散。这表明RNA和某些蛋白质分子起到稳定类核的作用。在*E.coli*中已发现若干种DNA结合蛋白(表2-2)。

**表2-2　大肠杆菌内的若干DNA结合蛋白**

| 蛋白质 | 组　　成 | 每个细胞含量 | 真核生物的相关蛋白质 | 基因位点 |
| --- | --- | --- | --- | --- |
| H | 两个28 KDa的相同亚基 | 30 000个二聚体 | 组蛋白　$H_2A$ | 未　知 |
| HU | 两个各9 KDa的不相同亚基 | 40 000个二聚体 | 组蛋白　$H_2B$ | 未　知 |
| $HLP_1$ | 17 KDa的单位 | 20 000个单体 | 未　知 | *firA* |
| P | 3 KDa的亚基 | 未　知 | 精蛋白 | 未　知 |

这些DNA结合蛋白质，使长达$4.2\times10^6$bp的*E.coli*染色体DNA压缩成为一个手脚架形(scaffold)结构。在这一结构的中心为多种DNA结合蛋白质，而DNA双螺旋分子有许多位点

与这些蛋白质结合，而形成大约 100 个小区(domain)。每个小区的 DNA 都是负超螺旋。一个小区的 DNA 有两个端点被蛋白质所固定，因而保持每个小区的相对独立性。当用极微量的DNA酶 I 作用于这个手脚架时，只能使少量的小区的 DNA 成为松弛状态，而其他小区仍然保持超螺旋状态，这就说明 *E.coli* 染色体 DNA 并非一个自由的 DNA 分子。这种手脚架形结构示于图 2-2。

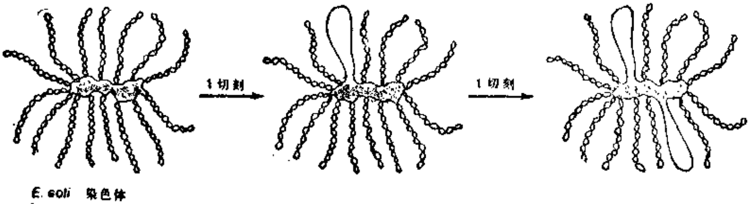
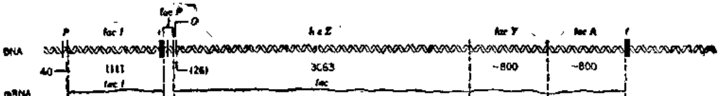



图 2-2　大肠杆菌染色体的手脚架形结构以及微量 DNAase I 作用示意图

大肠杆菌染色体是由 $4.2\times10^6$ 碱基对组成的双链环状 DNA 分子，以 1 000 bp 作为基因的平均长度计算，大肠杆菌可以有 3 000～4 000 个基因。目前已经定位的基因已达 900 多个。关于大肠杆菌的基因结构不能在此详述，但是我们可以从中总结出一些规律，表明生物在进化中选择了经济而又有效的结构形式：

(1) 功能上相关的几个结构基因前后相连，再加上一个共同的调节基因和一组共同的控制位点即启动子(promoter)和操作子 (operator)* 在基因转录时协同动作。对基因表达过程实行原发调节控制，也是最有效的一种调节控制。细菌基因表达调控的这样一个完整的单元，称为操纵元(operon)**。例如，大肠杆菌中司乳糖代谢的 β-半乳糖苷酶，半乳糖苷透性酶，半乳糖苷乙酰化酶等三个酶的基因 $z$、$y$、$a$ 与控制位点 $o$、$p$ 以及它们的调节基因 $i$ 一起组成乳糖操纵元(图 2-3)。




| | lac I | lac Z | lac Y | lac A |
|---|---|---|---|---|
| 多肽氨基酸数和分子量 | 360<br>38,000 | 1021<br>125,000 | ~275<br>30,000 | ~275<br>30,000 |
| 活性蛋白质和分子量 | 四聚体<br>152,000 | 四聚体<br>500,000 | 膜成分单体<br>30,000 | 二聚体<br>60 000 |
| 功能 | 阻遏蛋白 | β 半乳糖苷酶 | 透性酶 | 转乙酰酶 |

图 2-3　大肠杆菌乳糖操纵元及其 mRNA 和蛋白质

* OPERATOR 旧译“操纵基因”，然而“operator gene”一语在国际上自 70 年代中期已很少使用，至80年代几乎绝迹。Watson 的《Molecular Biology of the Gene》(1976,1987)，Gardner 的《Principles of Genetics》(1984)，Lewin 的《Genes (1985) 等著名教科书均取消“operator gene”一语，只用 operator，他们都指出 operator 只是一个阻遏蛋白质结合位点，最多 30～40 个 bp，并无基因特征，亦无基因产物。然而“操纵基因”一语在国内一直没有得到纠正。其实 operator 与 promoter，enhancer 一样，都是基因调控位点，因此本书使用“操作子”这一译名。

** OPERON，旧译"操纵子"。然而，operon 就像 neuron（神经元）是神经系统的结构和功能单位一样，它是基因表达和调控的结构上和功能上的单位。定义 operon 所使用的中心词是 unit（单元），是 DNA 上数以千计（kb）的核苷酸序列。它与启动子（promoter），终止子（terminator），衰减子（attenuator），增强子（enhancer），操作子（operator）等名词的概念完全不同。定义后者所使用的中心词是 site（位点），一般只有数十个 bp，多则百余 bp。同时，英文后缀 -on 有三种用法。在表示 DNA 上一段序列的词汇中的-on，除了 codon（密码子）外，并非来自 19 世纪法拉第所铸造的 ion 以及后来出现的 electron，proton，neutron 等词（见 Webster's Dictionary, College Edition）。从语义上来说，以-on 为后缀的表示 DNA 上一段序列的名词（codon 除外）只表示某种（某些）作用或过程所涉及的范围，operon 是如此，regulon，replicon，transcripton 等均是如此。例如，大肠杆菌（染色体）的 replicon 就是整个的 $4.2\times10^6$bp 的 DNA 分子。可是，以-or 或 -er 为后缀的表示 DNA 上一段序列的名词（如上面所列举的例子），是某种动作（作用或过程）的直接执行者之一（一般与一种或几种蛋白质或 RNA 共同完成这一动作）。

综上所述，英文中用以表示 DNA 上一段序列的名词有两类结尾；相应地，中译名也应该有两类结尾。根据以往的名词翻译先例，以 -on 为后缀的可翻译为"元"，以 -or 或 -er 为后缀的可翻译为"子"。这样，上述的 operon，regulon，replicon，transcripton 应译为操纵元，调控元，复制元，转录元；operator 就应该译为操作子。

最后，让我们认真地理解 Benjamin Lewin 在《Genes》第二版中对操纵元（operon）和操作子（operator）所作的定义：

OPERON is a complete unit of bacterial gene expression and regulation, including structural genes, regulator gene(s), and control elements in DNA recognized by regulator gene product(s).

OPERATOR is the site on DNA at which a repressor protein binds to prevent transcription from initiating at the adjacent promoter.

本书将在附录中列出涵义和译名有所变化的部份分子遗传学术语的英汉对照，并作简明的解释。凡本书作者所拟定的或建议的中译名将与流行的但不太妥当的中译名一并列出。

有时一个调节基因的产物能够调控几个操纵元中基因的表达。这样的几个操纵元加上它们的调节基因就构成了更大的基因表达调控单元，称为调控元（regulon）。其中的调节基因当然就不属于某一个操纵元，而是属于整个调控元。即使只调控一个操纵元的调节基因，在大多数情况下也并不与操纵元的主体部分相连。细胞内的蛋白质是可以到处扩散的，因而也没有必要与它控制的基因相毗邻。这样，在有些教科书中，则不把调节基因算作操纵元的成分，而是作为调控元的成分。

这些功能相关的位于一个操纵元内的若干个蛋白质基因在转录时产生一个多基因的 mRNA。有些功能相关的 RNA 基因也串联在一起，如构成核糖核蛋白体的三种 RNA 基因转录在同一个转录产物中，它们依次是：16S RNA…23S RNA…5S RNA。这三种 RNA 除了组建核糖体外，别无他用，而在核糖体中的比例又是 1:1:1。倘若它们不在同一个转录产物中，要末造成这三种 RNA 比例失调，影响细胞功能或造成浪费；要末需要一个极其复杂的耗费巨大的调节机构来保持正常的比例 1:1:1。当然，并非所有的功能相关的基因都组织在一起。

（2）蛋白质基因通常以单拷贝的形式存在。多拷贝基因或多拷贝序列存在于同一染色体上常常引起非均等交换，结果导致了相同序列之间的基因的缺失或倒位（详见第十一章）。然而，基因序列的倍增又是经常发生的（$10^{-4}$），可是绝大多数的倍增并没有给细菌带来选择上的优势，因而很快又被淘汰。这样，就使绝大多数的蛋白质基因保持单拷贝形式。

（3）RNA 基因通常是多拷贝的。大多数 *E.coli* 菌株都含有 16 S rRNA，23 S rRNA 和 5 S rRNA 基因各 7 个拷贝，以便装配大量的核糖体，使细菌能在 20 分钟内使细胞内的蛋白质含量加倍。在大肠杆菌中，rRNA基因除了多拷贝的特点外，它们在染色体上的分布也增加了

潜在的基因剂量。在7个rrn操纵元中，有6个分布在*E.coli* DNA的双向复制起点附近（遗传图上的83分），而不是在复制终点附近（遗传图上的33分）。可以设想，在一个细胞周期中，复制起点处的基因的表达量几乎相当于处于复制终点的同样基因的两倍。*E.coli* 7个rrn操纵元在染色体上的分布情况示于图2-4。

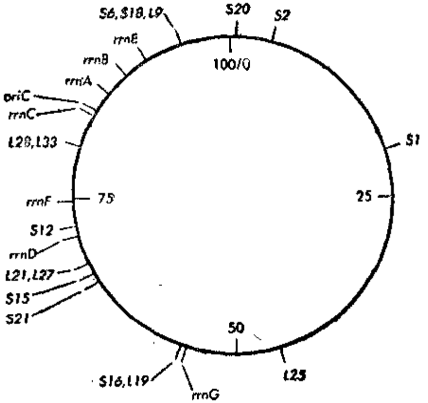



图2-4 *E.coli* rRNA基因以及核糖体蛋白质基因在染色体上的分布情况。rrn表示七个rRNA基因的操纵元，S和L分别表示小亚基和大亚基的操纵元，OriC为*E.coli*的双向复制起点。

(4) 除了把相关基因组织成操纵元的形式之外，各种基因的启动子和操作子部分的DNA序列也是多种多样的，以便与RNA聚合酶以及与调节基因的产物阻遏蛋白发生不同程度的结合，以对各种基因的表达加以精细的调节。

## 二、噬菌体

这里主要介绍大肠杆菌噬菌体$\phi\times$174和噬菌体$\lambda$。

$\phi\times174$是一种很小的大肠杆菌噬菌体，只有T类噬菌体大小的1/4，呈20面体颗粒状，没有尾巴（因而其DNA进入寄主细胞不是以注射方式进行）。其DNA为单链环状，共有5 386个核苷酸。其顺序于1977年由英国剑桥大学的Frederick Sanger分析确定。为此，Sanger第二次获得诺贝尔奖金。

$\phi\times174$共有11个基因，依次为A, A*, B, C, D, E, J, F, G, H和K（见图2-5和表2-3）。

由于$\phi\times174$的DNA数量有限，因而其基因在排列上更加体现了经济原则：

(1) $\phi\times174$有11个蛋白质基因，但是只转录成3个mRNA，其中一个从A基因开始，一个从B基因开始，另一个从D基因开始。

功能上相关的基因组织在一起并转录在同一个mRNA之中，这在原核生物中是很普遍的现象。

(2) $\phi\times174$的DNA分子绝大部分用来编码蛋白质，不翻译出来的部分只占4%(217/5386)，其中包括基因之间的间隔区和一些控制基因表达的序列。

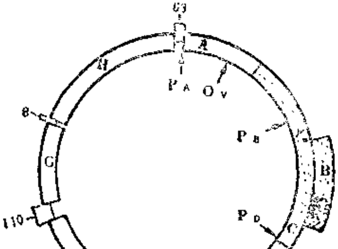



图2-5 $\phi\times174$的遗传图（数字表示不翻译的间隔区）

(3) $\phi\times174$的基因排列上最显著的特点是有重迭基因和基因内基因，这是通过对$\phi\times174$的DNA序列和它的9个基因产物的氨基酸序列进行了认真的比较工作而得出来，也是首次发现基因重叠现象。迄今为止，只有在噬菌体和

表 2-3 噬菌体 φ×174 的基因与功能

| 基因 | 功能 | 基因 | 功能 |
|---|---|---|---|
| A | RF 的复制，病毒链(正链)的合成 | J | 先结合于原头外的 DNA，然后在头内使 DNA 维持一定的结构 |
| A* | 关闭寄主 DNA 的合成 | | |
| B | 病毒外壳的形成 | F | 主要的外壳蛋白 |
| C | 将新合成的病毒链引入"原头"(prohead) | G | 大钉蛋白 |
| D | 病毒外壳的形成 | H | 小钉蛋白，能吸附于寄主细胞表面 |
| E | 细菌的裂解 | K | 不清楚 |

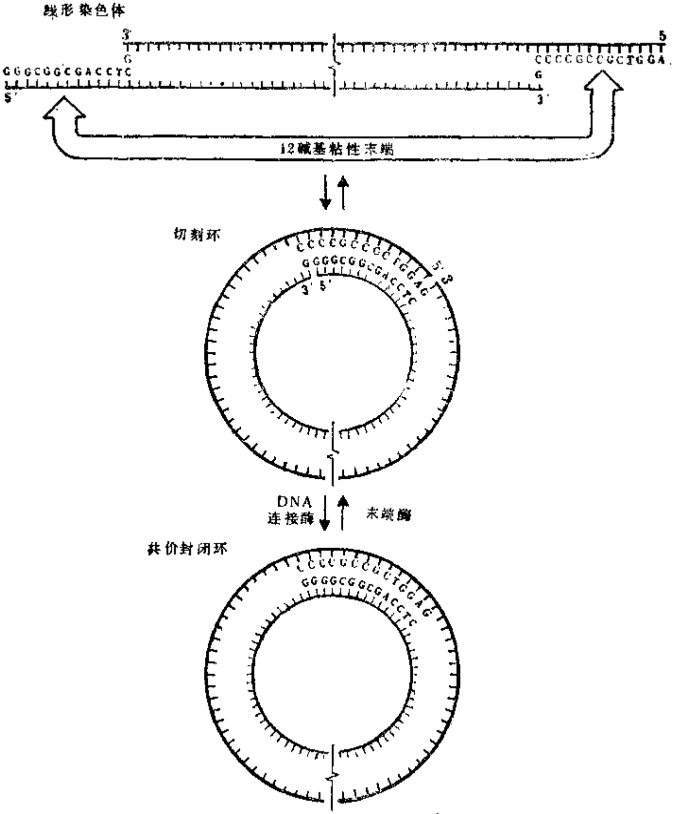



图 2-6 λ噬菌体 DNA 分子形式的转变

病毒中发现基因重叠现象。这可能是由于它们的 DNA 分子小，然而又必须有一定种类的基因产物。在进化过程中，这种压力导致了基因的重叠。

λ噬菌体的染色体共有 48 502 碱基对。分子量为 $3.2\times10^7$ 道尔顿，长度约 16 μm，被包装在由蛋白质组成的头部外壳之中。λ噬菌体的双链 DNA 有两种形式。一种是带有切刻的环状分子，切刻部分通过粘性末端的互补碱基之间的氢键而连接；另一种形式是两个粘性末端相互分离，形成线形分子。粘性末端有 12 个核苷酸，非共价结合的环状分子在加热时(70℃)很容易转变成线形分子；还可以用 DNA 连接酶将非共价结合的环状分子变为共价结合的封闭环状分子(图 2-6)。

对于λ噬菌体的遗传图，已经有相当深入的研究，见图 2-7。

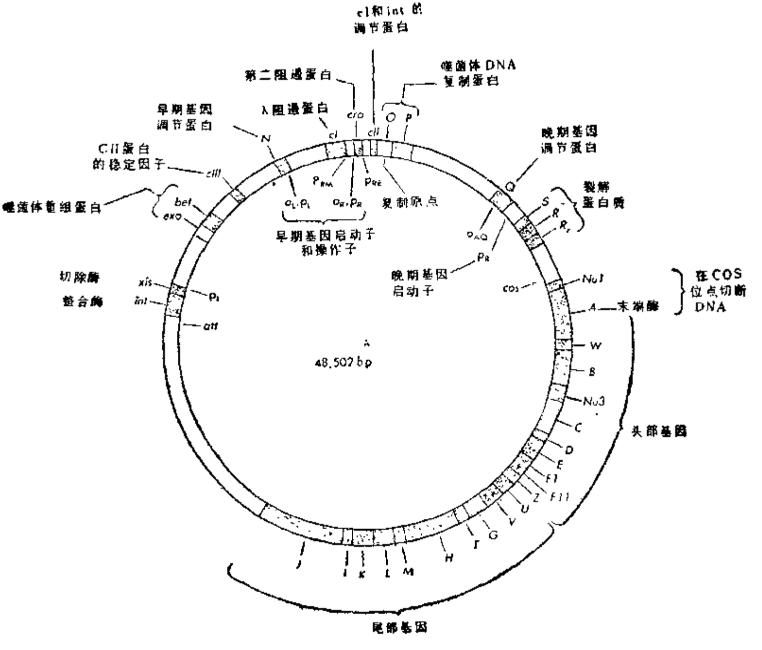



图 2-7 细胞内环形 λDNA 的基因图和主要的调控位点以及与 λDNA 基本形状变化(环形，线形和整合状态)有关的两个位点：*cos* 和 *att*。

λ噬菌体在大肠杆菌体内可以呈环形分子存在于细胞质中，也可以通过整合酶的作用而整合到寄主染色体上成为原噬菌体状态，并与寄主体染色体一起复制。这种状态能维持许多代，这种现象称为溶原化。溶原化的细胞之所以能维持溶原状态，就是由于整合在细菌染色体上的λ噬菌体基因中的 *cI* 基因能产生一种分子量为 52 000 道尔顿的阻遏蛋白(二聚体)。

这种蛋白能和左、右两个操作子相结合，阻碍两个早期 mRNA 的转录。因而，也就不能合成 λ 噬菌体复制所必需的蛋白质。这种阻遏蛋白不但能阻止原噬菌体本身 DNA 的复制，同样也能结合于新侵入的 λ 噬菌体的操作子，从而阻止它们的复制，使处于溶原状态的大肠杆菌不再被其他 λ 噬菌体的感染而裂解。也就是说，溶原性细胞获得了一种特殊的免疫能力，因而这种阻遏蛋白又称为免疫性物质。

然而，当 λ 噬菌体的 DNA 合成处于抑制状态时，某些因素(如紫外线照射，丝裂霉素 C 的处理)常会使阻遏蛋白失活，从而诱导原噬菌体发育为 λ 噬菌体而使细胞裂解。现在已经知道，"*cro*"基因产物能关闭阻遏蛋白的继续合成。

## 第四节　真核生物的染色体[9~27]

在真核生物的染色体中，DNA 约占 27%，组蛋白和非组蛋白占 66%，RNA 占 6%。

过去认为一条染色体可能是由若干段 DNA 构成的，间隔是蛋白质，但是随着 DNA 提取技术的改进，现在知道一个染色体(单体)中的 DNA 只是一个 DNA 分子。例如酵母菌在单倍体时期，17 条染色体就是 17 个 DNA 分子，人类染色体平均有 $1.27\times10^8$ 碱基对，平均分子量为 $8.38\times10^{10}$ 道尔顿。这就是说，每个体细胞的细胞核内含有 DNA 6.4 pg (picogram)。

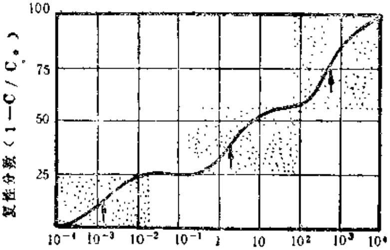
### 一、真核生物 DNA 复性动力学

从图 1-10 可以看出，各种原核生物 DNA 复性的 Cot 曲线形状都很相似，跨度一般只有两个数量级，只是 $Cot_{\frac{1}{2}}$ 值各不相同。也就是说，各种原核生物的 DNA 都是单一序列，只是其序列复杂性(complexity)各不相同而已。两种不同大小复杂性的 DNA 序列，当 DNA 绝对含量（即总的核苷酸数）相同时，则复杂性小的 DNA 分子浓度高，复性就快，需要的时间 $t$ 就短，因而 $Cot_{\frac{1}{2}}$ 值就小；不言而喻，复杂性大的 DNA 复性；所需的时间就长，$Cot_{\frac{1}{2}}$ 值就大。以大肠杆菌的 $Cot_{\frac{1}{2}}$ 为标准，就可以计算出待测样品 DNA 的复杂性(又称复杂长度)等于

$$4.2\times10^6\ \text{bp}\times\frac{\text{样品 DNA } Cot_{\frac{1}{2}}}{\text{大肠杆菌 DNA} Cot_{\frac{1}{2}}}$$

真核生物的 DNA 复性曲线就大不相同了，不但曲线形状不同，而且常跨越 7～8 个数量级。图 2-8 给出一个真核生物 DNA 复性曲线的模式图，实际图形要复杂得多，需要用计算机处理。如图 2-8 所示，复性反应分为三相，每相各代表基因组中不同复杂长度的序列类型，整个基因组 DNA 为 $7.0\times10^8$ bp。第一相称为快复性组分，代表了 25% 的总 DNA，Cot 值范围为 $10^{-4}\sim2\times10^{-2}$，$Cot_{\frac{1}{2}}=0.0013$。第二相称为中间复性组分，代表了 30% 的总 DNA，Cot 值的范围为 0.2～100，$Cot_{\frac{1}{2}}=1.9$。第三相为慢复性组分，占总 DNA 的 45%，Cot 值的范围为 80～10 000，$Cot_{\frac{1}{2}}=630$。

为了计算上述三种组分的复杂性，必须将每一组分当成是独立的组分以便与标准 DNA 的复性动力学曲线进行比较。对于第三组分来说，由于它占总 DNA 的45%，其浓度 Co 也是总Co




| | 快复性组分 | 中间复性组分 | 慢复性组分 |
|---|---|---|---|
| 占基因组百分数 | 25 | 30 | 45 |
| Cot 1/2 | 0.0013 | 1.9 | 630 |
| 复杂性(bp) | 340 | $6.0\times10^5$ | $3.0\times10^8$ |
| 重复频率 | 500,000 | 350 | 1 |

图 2-8　真核生物 DAN 复性动力学模式图

的 45%，因而它实际上的 $Cot_{\frac{1}{2}}$ 应该为 630×45%＝283，也就是说，如果把这一组分提取出来单独做复性试验，其 $cot_{\frac{1}{2}}$＝283，假定在同样条件下*E.coli* DNA的 $Cot_{\frac{1}{2}}$＝4.0，我们可算出第三组分的复杂性为 $3.0\times10^8$bp，同样可以算出第二组分为 $6\times10^5$ bp，第一组分为 340bp。

已知化学方法测量的 DNA 总长度为 $7.0\times10^8$bp。这一长度称为化学复杂长度。对于第三组分来说，其化学复杂长度为 $7.0\times10^8\times45\%=3.15\times10^8$bp，按照复性动力学计算出来的复杂长度为 $3.0\times10^8$，称为动力学复杂长度，二者数值上是很接近的。实际工作中，常常并不知道基因组的复杂长度（化学复杂长度）。这时可以根据单一序列（一般把基因组中只有2～3个拷贝的序列也算作单一序列）的动力学复杂长度计算出整个基因组的动力学复杂长度。比如上述例子中单一序列的动力学复杂长度为 $3.0\times10^8$ bp，则基因组动力学复杂长度为 $3.0\times10^8$ bp÷45%＝$6.7\times10^8$ bp，与其化学复杂长度值是很相近的。

根据基因组的总长度和每一组分的复杂长度，即可计算出每一组分的拷贝数，即重复频率(repetition frequency)，其定义可用如下公式来概括：

$$\text{重复频率 f} = \frac{\text{化学复杂长度}}{\text{动力学复杂长度}}$$

用此公式不难算出第三组分 f 值为 1，第二组分 f 值为 350，第一组分 f 值为 500 000。由于在同一个复性曲线上的各动力学组分的 $Cot_{\frac{1}{2}}$ 值与其重复频率成反比。即 $Cot_{\frac{1}{2}}(1):Cot_{\frac{1}{2}}(2)=f(2):f(1)$，式中(1)和(2)代表两个不同的动力学组分，f 代表其重复频率。由上式可得到

$$f(2)=\frac{Cot_{\frac{1}{2}}(1)}{Cot_{\frac{1}{2}}(2)}\cdot f(1)$$

若(1)为非重复序列，则 f(1)＝1，则可简化为 $f(2)=Cot_{\frac{1}{2}}(1)/Cot_{\frac{1}{2}}(2)$，用这一公式计算图2-8所给出的三种组分的重复频率：630÷1.9＝332（第二组分）；630÷0.0013＝485 000（第一组

分）。这种算法得出的值偏低，其原因在于非重复序列（单一序列）的重复频率并非为1.0，因为通常把2～3个拷贝的序列也算作非重复序列。在上例中其非重复序列的重复频率实际上为1.05。按此计算：$630 \div 1.9 \times 1.05 = 348$（第二组分）；$630 \div 0.0013 \times 1.05 = 509\,000$（第一组分），这样就与按定义由化学复杂长度除以动力学复杂长度而求出的重复频率很接近。

## 二、真核生物染色体上的单一序列和重复序列以及卫星 DNA

真核生物的 DNA 比原核生物大得多，DNA 序列的组织性也就复杂得多。根据 DNA 复性动力学的研究，DNA 序列可以分为四种类型：

1．*单一序列又称非重复序列*　在一个基因组中只有一个拷贝，真核生物的大多数基因在单倍体中都是单拷贝的。

2，*轻度重复序列*　在一个基因组中有2～10个拷贝（但是2～3拷贝常常被看作是非重复序列），如组蛋白基因和酵母 tRNA 基因。在某些情况下，这些序列并非一点不差地重复，这些基因的产物在氨基酸组成上略有差异，如胚胎、胎儿、成人的血红蛋白 β 链的一系列基因就是如此。

3．*中度重复序列*　10～几百个拷贝。中度重复序列一般是不编码的序列，据认为在基因调控中起重要的作用，包括开启或关闭基因的活性，促进或终止转录，DNA 复制的起始，其转录产物参与 hnRNA 的处理等。这种重复序列的平均长度大约300个 bp，常有数千种像这样的重复序列，每一种平均重复数百次。它们在一起构成了序列家族。它们与非重复序列相间排列。其中最使人感兴趣的就是人类的 Alu 序列家族（Alu family）和其他哺乳动物中的Alu相当序列家族（Alu—equivalent family），在小鼠中有500 000个与 Alu 序列相关的序列，称为 $B_1$ 序列家族。

Alu 族序列成员众多，大约有300 000个，（平均每6 kb DNA 就有一个）。每个长度约300bp，在其第170位置附近都有 AGCT 这样的序列，可被限制性内切酶 Alu I 所切割（AG↓CT），故此得名。无论从 Alu 族序列的长度还是从其重复频率上来看，Alu 族序列都更像高度重复序列；然而它们不同于高度重复序列的串联集中分布，而是广泛散布在非重复序列之间。Alu家族的各个成员之间有很大的同源性，个别成员与其一致序列（consensns sequence）之间有87％的同源性，每个序列的左边和右边都有相似的130bp序列，右边一半还插入一个31bp的无关序列。小鼠 $B_1$ 家族重复序列只有130bp，相当于 Alu 族重复序列的一个单体，$B_1$ 与 Alu 有70～80％的同源性。Alu 序列很像转座元，因为其两翼各有一个正向重复序列，然而在Alu 家族的各成员中这一正向重复序列的长度并不相同，这还是一个谜（见第十一章第九节）。Alu 序列中有一段14bp 的序列，与某些病毒（如 SV40）的复制起点附近的序列十分相似。然而 Alu 序列数要比预期的人类基因组的复制起点数目多10倍，目前尚无法解释。

由于 Alu 序列在人类基因组中的广泛存在，促使人们去探索其在基因调控中的功能

4．*高度重复序列*　几百个到几百万个拷贝。在高度重复序列中，有一些是重复数百次的基因，如 rRNA 基因和某些 tRNA 基因；大多数是重复程度更高的序列。例如在果蝇（*D. virilis*）染色体的着丝点附近就存在着几种高度重复序列：ACAAACT $1.1 \times 10^7$ 拷贝，占基

因组 DNA 的 25%，ATAAACT ACAAATT 各 $3.6\times10^6$ 拷贝，各占基因组 DNA 的 8%。每一种高度重复序列中，一条链含 T·G 较多，增加了它的浮力密度，因此在变性时成为重链，另一条成为轻链。这种两条链上的不对称性常常是许多高度重复序列的特征之一。高度重复序列一般不转录，因为在这些极其简单的核苷酸顺序上缺少转录所必需的启动子。然而，在 DNA 复制时它们却是和单一序列复制得一样快。大多数高等真核生物的 DNA 都会有 20% 以上的高度重复序列。

在高度重复序列中，常有一些 A·T 含量很高的简单高度重复序列，例如螃蟹DNA 含有 AT 的简单重复序列，有时在 30 个左右的碱基对中才偶尔插入一个 G·C 对，因而 A·T 含量达到 97%。由于 A·T 段浮力密度较小，因而在将 DNA 切断成数百个碱基对的片断进行超离心时，常会在主要的 DNA 带的上面有一个次要的 DNA 带相伴随，这就是所谓卫星 DNA(主带的 AT 含量为 58%)。有的高度重复序列的碱基组成与总体的碱基组成差异不大，因而并非所有的高度重复序列都能成为卫星 DNA。在这种情况下，可以用其他实验手段，如复性试验把高度重复的 DNA 分离出来并进行鉴定，可以发现他们与卫星 DNA 有相同的特点，即串联集中分布。有时把这种高度重复序列叫做隐蔽卫星 DNA(cryptic satellite DNA)。明显的卫星 DNA 与隐蔽的卫星 DNA 加在一起通常可以包括基因组中高度重复DNA的几乎全部串联重复区段。

卫星 DNA 在染色体上的位置可以用分子原位杂交 (in situ hybridization 或 cytological hybridization)方法进行鉴定。其方法是在盖玻片上将细胞裂解，变性，再滴加放射性标记的单链(卫星DNA或其RNA)，适当洗涤，然后再通过放射自显影方法得到杂交情况的照片。实验表明，卫星DNA位于着丝点附近的异染色质区。人们很早就知道异染色质区的 DNA 是不表达的，不产生 RNA 和蛋白质产物。细胞内也确实不存在于与卫星 DNA 序列相应的 RNA。卫星 DNA 在着丝点处的集中分布可能与细胞分裂时染色体的运动有关。

### 三、卫星 DNA 的等级结构及其起源和进化

哺乳动物的卫星 DNA 常常是多等级的，也就是说一个大的重复单位是由若干个彼此相似的小重复单位串联组成的；每个小重复单位是由若干个彼此相似的更小的重复单位串联组成的。例如将小鼠卫星 DNA 用限制酶 EcoRⅡ (识别位点为 ↓CC($\frac{A}{T}$)GG)进行切割，可以得到 230～240 bp 的片断，通过序列分析发现一个 234bp 序列(图 2-9)，60～70% 的卫星 DNA 都含有这样的序列。如果把 234 bp 的左一半和右一半各 117 bp 排起来，发现只有 22 个 bp 不同，差异为 19%。这说明现在的 234 bp 重复单位是由 117 bp 亚重复单位倍生(duplication)而来，然后由于突变而积累了差异。117bp 的亚位单又可分成两个 58bp 的四分之一亚单位。四个 58bp 亚单位之间的差异达到 40%。58 bp 的亚单位又可以分成两个八分之一的亚单位，一是 28bp($\alpha$ 亚单位)，一是 30bp($\beta$ 亚单位)。八个八分之一亚单位之间的差异达 61%。如果再将八分之一亚单位分成三个部分，则每一部分都与如下一致顺序相近似(图 2-9)。

GAAAAACGT
GAAAAATGA
GAAAAAACT

小鼠的 234 bp 重复 DNA 序列可能是起源于某一个 9 bp 的序列(例如可能是

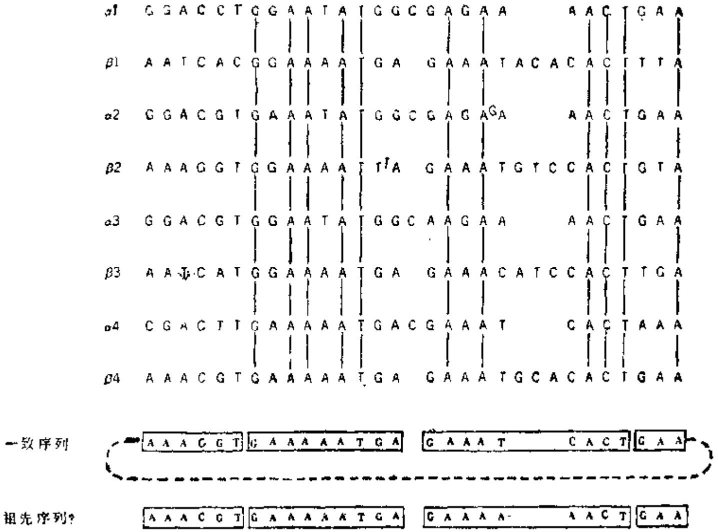



图 2-9 小鼠卫星 DNA 的 234bp 重复单位及其八分之一亚单位(四个α亚单位和四个β亚单位)。三个框内表示三种一致序列以及可能的祖先序列

GAAAAATGT),其演化原理是:在某一特定时刻可能由于某种原因使一个 DNA 序列突然横向扩增,产生多个串联重复单位。这种突然的横向扩增就称为跃进式复制 (saltatory replication。随后,由于突变的积累,使各重复单位失去了同一性。然后再经历一次跃进式复制,出现更长的重复单位;由于再经过突变的积累,使这更长的重复单位又失去同一性。如此反复,直至今日,以后仍可能继续进行下去。这样,9→27→58→117→234 经过了四轮跃进式复制和突变积累过程,在由 27→58 的过程中,还包括了两个插入突变,一个插入 C 成 α 亚单位,一个插入 3 个碱基成为 β 亚单位(图 2-10)。现在的 234 bp 的重复序列也并不完全相同,已经有突变积累,有的 EcoR Ⅱ 识别位点发生突变,以致于用 EcoR Ⅱ 切割时产生 468 bp 的二聚体和更大的多聚体(见图 2-13)。

高度重复 DNA 拷贝数的扩增或减少可能是由 DNA 的不均等交换(unequal crossover)引起的。细胞分裂时,染色体的同源重组是十分严格的,只能在同源序列之间进行。然而,由于卫星 DNA 的高度串联重复,一个重复序列很可能和另一条 DNA 分子上不等位的相同重复序列配对并发生交换。例如在两条染色体上各有 5 个 ab 串联重复单位,如果按图 2-11 发生非等位配对并发生交换时,则可得到一条含 7 个 ab 重复单位,而另一条只含 3 个 ab 重复单位。

图 2-11 所示的不均等交换是全符合状态的 (in register) 即某个染色体上的重复单位与

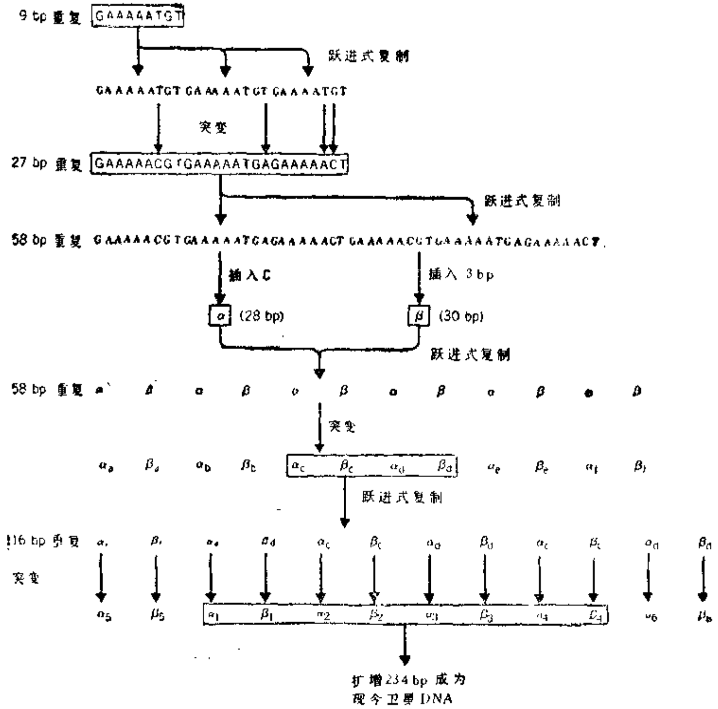



图 2-10 小鼠卫星 DNA 通过跃进式复制和突变积累的交替过程而演化

XabababababY
×
XabababababY
→
XababababababababY
XabababY

图 2-11 DNA 的不均等交换(全符合状态)

另一条染色体上不等位的整个重复单位完全相配而发生不均等交换。如果在一个重复单位中的两个部分,如图 2-11 中的 a 和 b,又有很大的同源性,则 a 和 b 也能发生配对而出现不均等交换。这种非等位配对是半符合状态 (half-register)。这种不均等交换的结果如图 2-12 所示。

XabababababY
×
XabababababY
→
XabababababbababY
XababaabY

图 2-12 半符合状态的不均等交换

半符合状态的不均等交换就解释了下述实验结果(图 2-13)。用限制酶切割卫星 DNA 时，除了产生单体重复单位和多聚体重复单位这样的整数倍重复单位以外，还产生了 1/2 倍，$1^1/_2$ 倍，$2^1/_2$ 倍等非整数倍的重复单位。在图 2-13 中，如果提高分辨率，还会发现 1/4 和 3/4 等非整数重复单位。

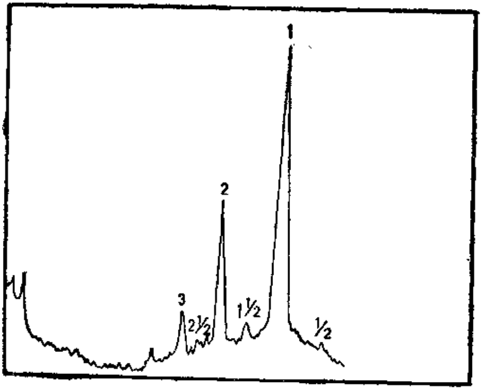


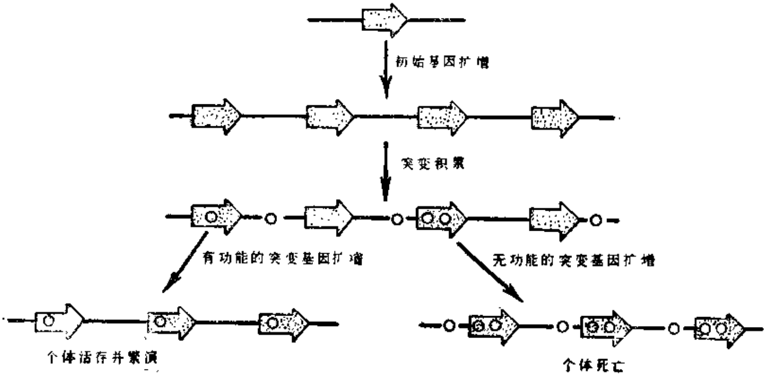

图 2-13 小鼠卫星 DNA 的 EcoR Ⅱ 切割的电泳扫描图

一般说来，卫星 DNA 不受任何选择压力，因为它不编码蛋白质或 RNA；其功能或许是其存在而不是其序列。这正如上面我们所看到的小鼠卫星 DNA 那样的情况，其各个拷贝之间相关而不是相同。然而，有些卫星 DNA，如节肢动物的卫星 DNA 是由几乎相同的重复单位组成的。人类的 U2 snRNA 基因有 10～20 个拷贝，在第 17 号染色体上串联集中分布。每个重复单元约 6 000 bp，然而其编码序列却只有 188 bp。编码序列受到自然选择的压力而保持相同，这是不难理解的。难以理解的是其余的 5 800 bp 的间隔序列为什么也要保持相同。当然其中包含了一部分转录起始、终止以及转录后处理的信号，它们必须是保守的。但是这一部分序列也只有 100bp 左右，仍然无法解释其余 5 700 bp 的保守性。从上述两个例子中，人们不可避免地会得到下述结论：任何串联重复的 DNA 序列，不管其中是否含有编码的遗传信息，都将经受均一化作用(homogenization)。尽管这种均一化机制仍然众说纷云，但其存在却是公认的。使串联重复的 DNA 序列保持均一化，目前一般认为有两种机制：一是交换固定 (crossover fixation)，二是基因转换(gene conversion)。这两种机制与基因扩增一起维持了串联重复序列的均一化(图2-14)。

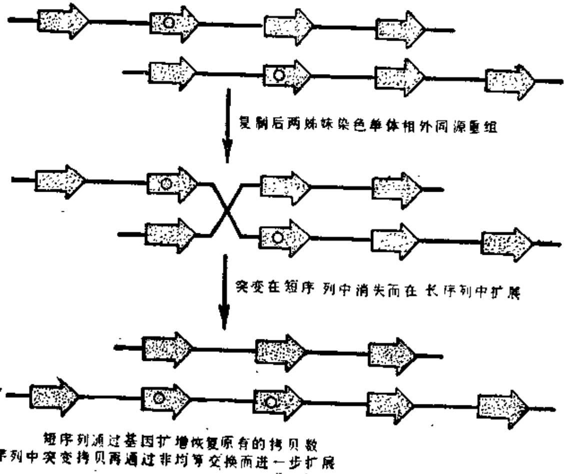
交换固定是由相外同源重组(homologous out-of-phase recombination)而引起的。也就是说在姊妹染色子体上的非等位重复序列之间发生了同源重组，其结果产生了非均等交换(图 2-14 b)。对于无编码功能的串联重复序列来说，非均等交换的两种产物均可能传递下去。因此突变序列要么被淘汰，要么迅速占据整个重复序列。而对于串联重复基因来说，在大多数情况下突变基因将被淘汰；但有功能的突变基因序列亦有可能迅速扩展。在非均等交换中拷贝数减少了的重复序列可以在后代中通过基因扩增恢复原有的拷贝数，例如果蝇的 rDNA 的拷贝数的变化。

基因转换也是通过同源序列之间的重组而实现的。同源重组的中间体经拆分以后形成异

(a) 基因扩增


初始基因扩增
突变积累
有功能的突变基因扩增
无功能的突变基因扩增
个体活存并繁演
个体死亡

(b) 交换固定


复制后两姊妹染色单体相外同源重组
突变在短序列中消失而在长序列中扩展
短序列通过基因扩增恢复原有的拷贝数
长序列中突变拷贝再通过非均等交换而进一步扩展

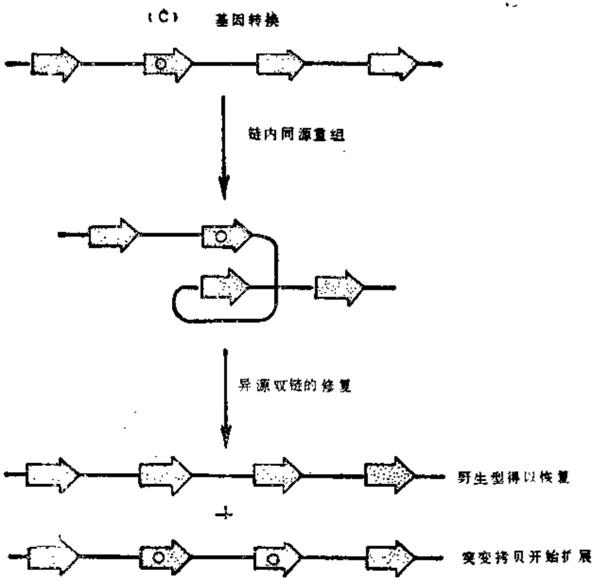



图 2-14 交换固定和基因转换与基因扩增共同维持了串联序列重复序列的均一性
(a) 基因扩增； (b) 交换固定； (c) 基因转换

源双链，再经修复后产生基因转换(详见第十一章)。基因转换的结果示于图 2-14c。需要指出的是，基因转换同样能使串联重复序列保持均一化。

### 四、染色质和核小体

**1. 染色质** 真核生物的染色体在细胞生活周期的大部分时间里都是以染色质(chromatin)的形式存在的。染色质是一种纤维状结构，叫做染色质丝，它是由最基本的单位——核小体(nucleosome)成串排列而成的。在细胞核的大部分区域，染色质丝的折叠压缩程度比较小，进行细胞染色时着色较浅，这一部分染色质叫做常染色质(euchromatin)，常染色质中DNA的包装比(packing ratio)约为1 000～2 000，即DNA的实际长度是染色质长度的1 000～2 000倍。细胞分裂时，常染色质进一步折叠压缩，形成典型的染色体结构(图 2-15)；这时DNA的包装比约为10 000。而着丝点等部位的染色质丝在细胞间期就折叠压缩得非常紧密，和细胞分裂时染色体的情况差不多。这部分叫做异染色质(heterochromatin)，其DNA包装比在整个细胞周期中基本上没有变化。异染色质又可分为两类：一类叫做组成型异染色质(constitutive heterochromatin)，其DNA序列从来不转录，如卫星DNA；另一类叫机动性异染色质(facultative heterochromatin)，它在某些情况下为常染色质，在某些情况下(不同性别，不同组织等)又可能成为异染色质。典型的例子就是哺乳动物雌性个体中的两个X染色体中有一个随机的失活，没有转录活性，这个作用称为异染色质化。遗传物质的紧密折叠压缩可能是关

闭基因活性的一种手段。那么，常染色质部分的基因是否都具有转录活性呢？也不是。在一个特定的时间内，只有少部分基因在转录。可见，处于常染色质状态只是基因转录的必要条件，而不是充分条件。

染色质和染色体的极大的包装比就提示了，DNA不可能直接包装到细微的染色质丝中去，更不可能直接构成染色体，而是必须有一系列的结构等级：DNA 和组蛋白构成核小体，核小体再绕成一个中空的螺线管状结构，这种螺线管状结构（有的部分就是珠状核小体结构）就成为染色质丝。染色质丝再与许多非组蛋白结合形成染色体结构。这种结构等级示于图 2-15。

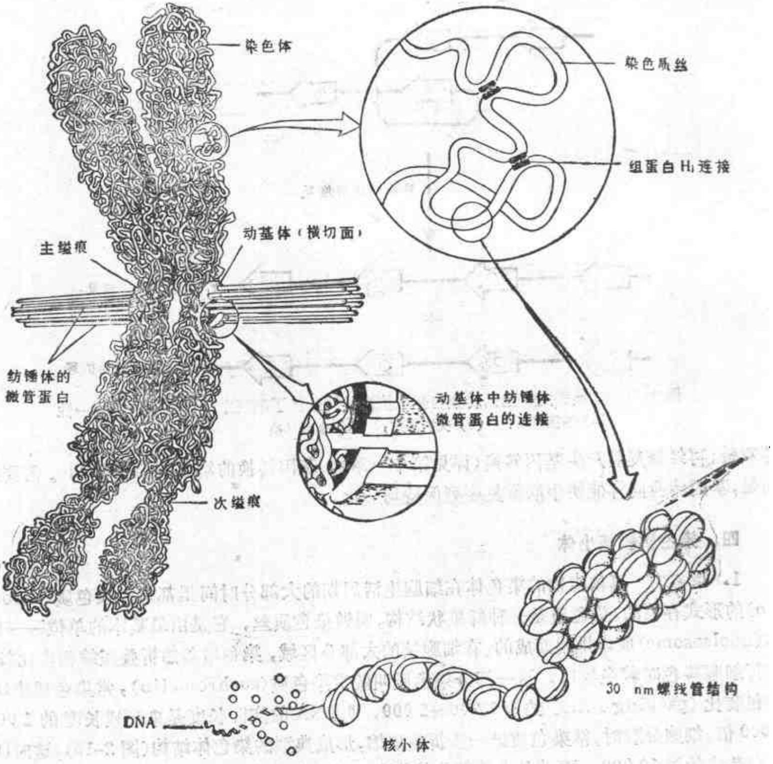



图 2-15 染色体的结构等级

2. 核小体 核小体是构成染色质的基本结构单位，使得染色质中 DNA、RNA 和蛋白质组织成为一种致密的结构形式。早在 1974 年人们就认识到所有真核生物染色质都具有同样类型的结构——核小体亚单位。如果用足量的微球菌核酸酶（micrococcal nuclease）处理染色质，则可以得核小体单体；如果用微量的酶作不彻底切割，则可以得到核小体的二体、三体等。

每个核小体单位包括 200bp 左右的 DNA 和一个组蛋白八聚体以及一个分子的组蛋白 H1。组蛋白八聚体核心颗粒是由H2A、H2B、H3 和 H4 各两个分子所组成，因而这四种组蛋白

又常称作核心组蛋白，位于组蛋白八聚体核心颗粒表面的DNA则称为核心DNA(core DNA)，而位于两个核心颗粒之间的DNA则称为连接DNA(linker DNA)、组蛋白H1恰恰就在连接DNA上。

组蛋白是一种碱性蛋白质，等电点一般在pH 10.0以上，是由于它们含有较多的碱性氨基酸的缘故。分子量一般为1万多至2万多。提取组蛋白，可以用0.2mol/L $H_2SO_4$，组蛋白可以溶解，而DNA却不溶解。

一般真核生物的染色体上有五种主要的组蛋白，分别名为H1、H2A、H2B、H3、H4，它们在同一生物的不同组织中全一样，在不同的真核生物中也很相似。

H 4：由102个氨基酸所组成，分子量11 282，富含精氨酸和甘氨酸。H 4在进化中极为保守，几乎每种真核生物都一样。

H3：135个氨基酸(包括二个半胱氨酸，但大鼠和豌豆只有一个半胱氨酸)分子量153 000，富含精氨酸。H3在进化中也很保守。

H2B：125个氨基酸，分子量13 700，稍富含赖氨酸，H2B在进化中也是保守的。

H2A：129个氨基酸，分子量14 500，稍富含赖氨酸，H2A在进化中也较为保守。

H1：215个氨基酸，分子量23 000，富含赖氨酸。在进化中H1最不保守。在同一种生物中，也有多种H1。

五种组蛋白质中，除H1的N端富含疏水氨基酸，C端富含碱性氨基酸以外，其余四种组蛋白都是N端富含碱性氨基酸(如精氨酸，赖氨酸)，C端富含疏水氨基酸(如缬氨酸、异亮氨酸)。五种组蛋白分子的基本参数列于表2-4。

表2-4 五种组蛋白分子的基本参数

| 种类 | 类型 | 碱性氨基酸 | | | 酸性氨基酸 | 碱性氨基酸/酸性氨基酸 | 氨基酸残基数 | 分子量 | 核小体上位置 |
|---|---|---|---|---|---|---|---|---|---|
| | | Lys | Arg | Lys/Arg | | | | | |
| H1 | 极度富含Lys | 29% | 1% | 29 | 5% | 5.4 | 215 | 23 000 | 连接* |
| H2A | 〃 | 11% | 9% | 1.2 | 15% | 1.4 | 129 | 14 500 | 核心 |
| H2B | 〃 | 16% | 6% | 2.7 | 13% | 1.7 | 125 | 13 774 | 〃 |
| H3 | 轻度富含Lys+富含Arg | 10% | 13% | 0.78 | 13% | 1.8 | 135 | 15 324 | 〃 |
| H4 | 〃 | 11% | 14% | 0.79 | 10% | 2.5 | 102 | 11 282 | 〃 |

* 现已证明：H1是与核小体的核心颗粒相当靠近的，认为它位于连接DNA上，只是一种习惯说法。

在五种组蛋白中，除了H1以外，其余四种有相互作用以形成聚合体的趋势，它们通过C端的疏水氨基酸相结合，而N端的碱性氨基酸则向四面伸出以便与DNA分子相互作用。

H3和H4的亲和力最强，H2A和H2B的亲和力也很强，H2B和H4也容易结合。四种组蛋白的亲和力可用图解说明如下(图2-16)。

四种组蛋白分子是如何结合成八聚体的呢？早期曾经认为是两个H3分子和两个H4分子形成四聚体，两个H2A分子和两个H2B分子结合成四聚体，然后两个四聚体再结合成八聚体。通过A.Klug等人用X衍射以及中子散射等方法对核小体结构的研究，认为是两个

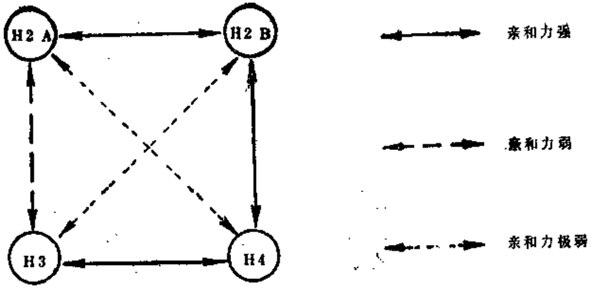



图 2-16 四种组蛋白亲和力图解

H3 分子和两个 H4 分子先形成四聚体，然后再与两个由 H2A 和 H2B 构成的异二聚体(heterodimer) 结合成为八聚体。现已知道，核小体具有二分对称性 (dyad symmetry)，四聚体 $(H3)_2(H4)_2$ 中的两个 H3 的连接处恰恰就在这一对称轴上。四种核心组蛋白分子除了其氨基末端的碱性氨基酸尾巴之外，都只有一个结构域(Single domain)，它们在八聚体上相互连接的次序是 H2A—H2B—H4—H3—H3—H4—H2B—H2A。这一连接顺序与上述的四种核心组蛋白分子的亲和力实验结果完全一致。八个核心组蛋白分子连接成为匝道(ramp)形状，直径为 70Å。螺距大约为 27Å。这样的结构恰恰提供了 1.75 圈 DNA(146 bp)缠绕的模板，使之成为所观察到的核小体尺寸：核小体颗粒高约 60Å，直径约 110Å，DNA 在核小体外绕 $1^3/_4$ 圈。组蛋白在核小体中的相对位置示于图 2-17。

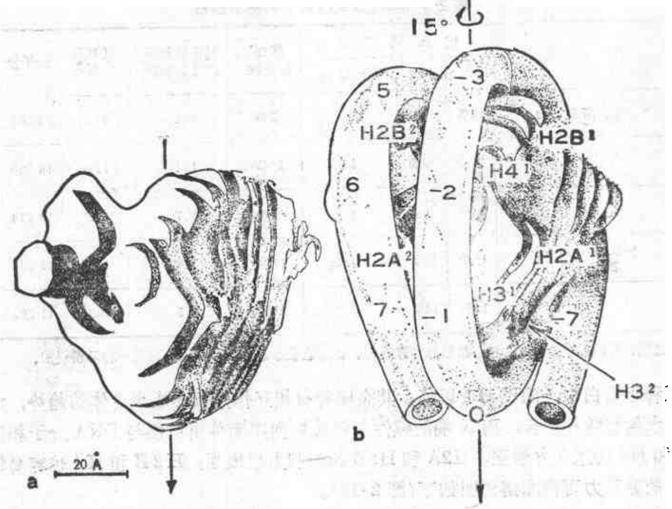



图 2-17 核小体中各组蛋白分子的位置。图中 −7～+7 表示 146 bp 的 DNA 上的位点，每位点相距大约 10 bp，↓表示二分对称的对称轴。
(a) 核心颗粒；(b) 核心颗粒外缠绕 $1^3/_4$ 圈 DNA

核小体组装的顺序是：H3 和 H4 先形成四聚体$(H3)_2(H4)_2$，这个四聚体像一个端点错开的弹簧垫圈一样，它决定了与之结合的核小体 DNA 中心节段大约 70～80 bp 的走向并由此决定了整个核小体 DNA 的走向；然后 H2A 和 H2B 形成的异二聚体结合于四聚体的两个侧面，每个异二聚体与中心节段 DNA 两侧各约 30～40bp 的序列结合；最后，组蛋白 H1 与特定的DNA 构象结合，封住了核小体的进口和出口(图2-18)。这一模型与 Felsonfeld 等人人实验结果相符，即在缺少 H2A 与 H2B 的情况下，H3、H4 也能够与 DNA 形成类似于核小体的结构，特别是表现在超螺旋以及抵抗微球菌核酸酶的降解等性质方面；而单独的 H2A、H2B 或 H2A—H2B 则无此能力。在这个模型中，H3 二聚体处于中心位置，对称轴穿过其连接处。根据 Klug等人的核酸酶降解试验，对称轴所穿过的大沟是朝向 H3二聚体，这与典型的 Cro 蛋白质，CAP 蛋白质以及λ阻遏蛋白等调节蛋白质与DNA序列的作用不同。此外，H3 分子第 110 位上的 cys 很可能相互结合成二硫桥(图 2-19)。

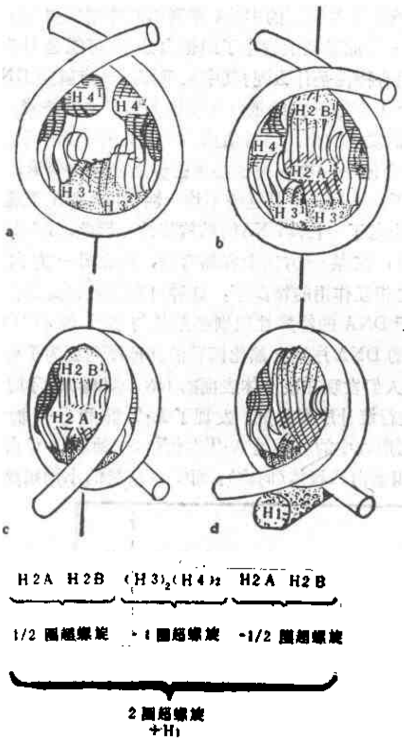



图 2-18　核小体组装过程示意图

(a) $(H_3)_2(H4)_2$ 四聚体形成一个锁蘸样结构，70～80 bp DNA 缠绕在上面，形成一圈超螺旋。(b) H2A—H2B 异二聚体结合于四聚体的两侧面。(c) 各个H2A～H2B异二聚体与 30～40bp 结合，各产生半圈超螺旋。(d) H1 与 DNA 结合，锁住核小体的进出口。

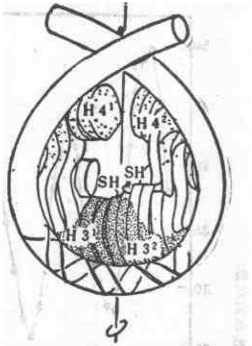



图 2-19　组蛋白四聚体与对称轴附近的两圈DNA螺旋，对称轴所通过的大沟朝向 H3 二聚体

关于核小体在 DNA 双螺旋链上定位(亦称定相)问题是近几年来染色质结构研究方面的焦点之一。这个问题不仅是核小体装配和染色质结构问题，而且与 DNA的复制和基因的转录密切相关。

鉴于组蛋白核心颗粒的二分对称状况，核心 DNA 是否也有序列的对称性或某种序列的优先性？Witting 等人(1982年)曾提出过 TTCGA 五核苷酸序列是核小体定位所必须的，但是实验结果远非如此。因为能够与组蛋白核心颗粒相连接的 DNA 序列具有极大的多样性。于

Q75
CNE

是人们又提出一个调和的解释,即核小体定位具有某种序列优先性(如上述的TTCGA),即以该序列为识别信号先在一段DNA上形成第一个核小体,然后按照距离形成其余的核小体。这样的调和的解释,无法说明核小体间DNA(即连接DNA)的长短不一。

由此看来,核小体定位主要不是由组蛋白八聚体与特定的DNA序列相互作用而决定的,而是由DNA双螺旋本身的固有特性所决定的;在此基础上发生了组蛋白分子(可能是H3—H3)与特定的DNA构象发生相互作用。这种固有特性是什么呢?其中最重要的一点就是DNA双螺旋的可弯曲性(bendability)。事实上,DNA双螺旋在组蛋白八聚体上紧密地缠绕着。这时,DNA双螺旋的两条沟(大沟和小沟)也必然发生变化:向着组蛋白八聚体的一面变窄,向外的一面变宽。在这样的DNA与组蛋白的复合体中,DNA分子必须能够适应这种变形,第一章中我们已经讲到DNA的碱基序列分布的不均一性。正是这些不均一性造成DNA双螺旋的局部结构特征;这些局部特征的加和效应就决定了一长段DNA的结构特性。因此,DNA双螺旋一般并不表现各向同性,而是各向异性的;在某一方向上容易弯曲,而在另一方向上可能相反。必须指出,即使在缺乏蛋白质与之相互作用的情况下,这种可弯曲性也是存在的(固有的内在的可弯曲性)。组蛋白八聚体对于DNA的特异性识别必须是与整个核心DNA相互作用的结果,而位于八聚体二分对称点处的DNA序列必然比两臂的其他序列更为重要。人们用DNAase I切割含有核小体的DNA,人们发现在核小体表面的DNA双螺旋周期是10.32bp/圈。人们通过对大量的核心DNA进行统计序列分析,发现了A·T群和G·C群按照≈10.2bp的周期进行调制;而且这两种调制形式中的一种是A·T占优势,一种是G·C占优势。进一步分析发现,A·T群处的小沟向着组蛋白八聚体(向内),而G·C群处的小沟则离开

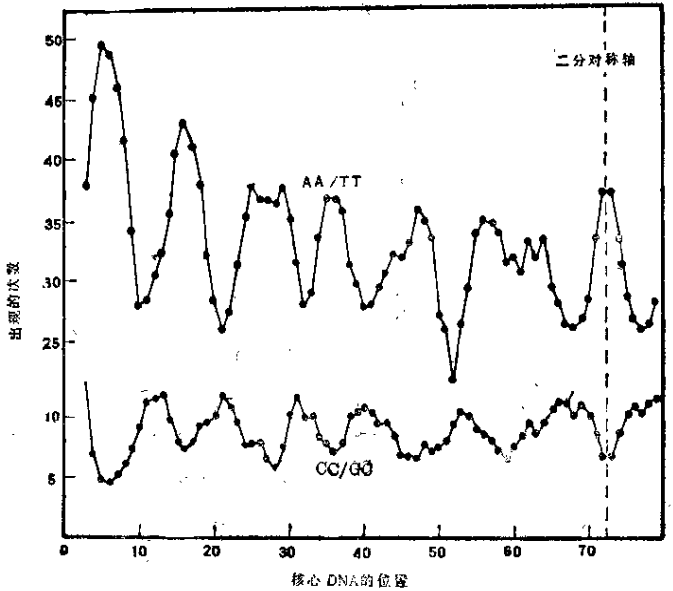



图2-20 AA/TT和GC/GC两种二核苷酸序列在177个核心DNA中出现的次数

八聚体向外。A.A Travors 等人(1986)对 177 个核心 DNA 序列进行详细的研究(图 2-20)，发现 A·T 群调制是以 A—A/T—T 二核苷酸形式出现的，这也包括 A—A—A/T—T—T 和 A—A—T/A—T—T 两种三核苷酸形式，但其他含有 A—A/T—T 的三核苷酸形式则不适合，如 C—A—A/T—T—G。而 G·C 群调制是以 G—C/G—C 二核苷酸形式出现的，这也包括 G—G—C/G—C—C 这样的三核苷酸形式，而其他包含 G—C/G—C 的三核苷酸形式则可能较小，如 A—G—C/G—C—T。

上面二核苷酸调制形式只是提供了DNA双螺旋可弯曲性的基础，单靠这一点还不能够使组蛋白八聚体精确定位，因为八聚体如果移动 10.2 个 bp 仍然符合调制的周期性。在八聚体的二分对称点处，小沟是离开核心颗粒向外的，而二分对称点又恰恰与 A—A/T—T 相结合，这里必然存在着一个相变问题。关于相交的细节目前还不清楚，这可能涉及 H3—H3 与这一节段 DNA 相互作用并有可能改变这一小段 DNA 的双螺旋周期。

关于组蛋白 H1 的位置和作用近年来也取得新的进展。H1 分子很小，无法在电子显微镜下看到；然而，可以通过它对部分折叠的染色质状态的影响来推测它的位置。在 1～5 mmol/L 的离子强度时，在电子显微镜下观察含有 H1 和不含 H1 的染色质，发现有 H1 的染色质呈现有序的结构，其中核小体规则地排列成之字路径，(其平的两面一面贴在金属网上，一面朝上)。之所以如此，是因为进入核小体的 DNA 和离开核小体的 DNA 靠得很近，因为核小体差不多有两整圈超螺旋。面缺少 H1 的染色质，进入核小体和离开核小体的 DNA 链差不多在核小体两侧，因而电子显微镜下呈直线形排列，而且这样的核小体容易散开。H1 的存在则防止其散开，H1 必定是稳定 DNA 进出核小体部位的结构，可能是基于特定 DNA 构象的识别。用微球菌核酸酶降解染色质，在反应的早期阶段，确实可以得到 166bp 的片断，但是很不稳定，进一步降解则得到比较稳定的 146bp。这可能是两圈 DNA 中每端各有 10 bp 从核心颗粒上翘起来，容易遭微球菌核酸酶的攻击。由于各 10bp 的 DNA 序列的两端相距很近，能同时与 H1 分子结合而被锁住。这样，由 166bp 的 DNA 绕在组蛋白八聚体外面形成二圈超螺旋并由 H1 在进出口将 DNA 锁住。这样的结构称之为染色质小体 (chromatosome)，其实这应该是更为正确的核小体的含义。人们所发现的各种生物染色质 DNA 的重复单位总是大于 160～170 bp，因此许多学者认为应该取 166～168 bp 而不是 146 bp 作为核小体结构的一个不变因素，而连接 DNA 则会相应地缩短，如小鼠染色质 DNA 重复单位为 200 bp，面连接 DNA 则为 32～34bp，有的生物的连接 DNA 可能近似于 0。

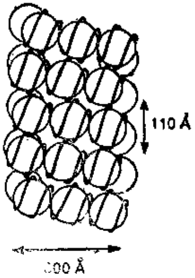


图 2-21　染色质的螺线管状结构，每圈螺旋由六个核小体组成

核小体与连接 DNA 构成串珠状的染色质丝。在有丝分裂过程中，串珠状的核蛋白纤丝进一步螺旋化和折迭成为直径约 300 Å 的螺线管结构，每圈螺旋有 6 个核小体组成。螺线管的中央有直径约 100 Å 的空腔(图 2-21)。H 1 组蛋白连接在核小体与连接 DNA 的会合点，不仅维持了核小体的稳定，而且对于维持这样高一级的结构负有主要责任。体外实验证明，H1 组蛋白能使DNA的结构更加致密。由于 H1 维持了染色质的高级结构，可以设想，在细胞周期中各个不同时相的染色体形态变化可能与 H1 组蛋白的数量和种类变化有密切关系。

实验证明，在 DNA 复制以后，新合成的组蛋白 H4、H3、H2A、H2B 将形成新的组蛋白八聚体，面原有的八聚体不再解

离。也就是说，在染色体的复制中，DNA 是半保守；而 DNA 所缠绕的组蛋白八聚体却是全保守。以后我们谈到 DNA 的复制机理时再讨论组蛋白八聚体的组装问题。

DNA 在核小体上的走向，决定于 $(H3)_2(H4)_2$ 四聚体的构象。DNA 在核小体上总是左手方向缠绕的，这与单独的 DNA 分子在溶液中自身缠绕成的超螺旋方向正好相反，然而却同是负超螺旋。这是同样的拓扑参数的不同表现，区别在于一个是自身缠绕，一个是支持物上缠绕。在核小体结构方面有不少的工作都是以 SV 40 的微染色体 (minichromosome)。SV 40 DNA 是 5 243 bp 的环形DNA（轮廓长度为1 500nm），在寄主体内能与寄主的组蛋白八聚体形成核小体结构，其轮廓长度为 210 nm。这样，包装比等于 1 500/210≈7。改变盐浓度，可以得到较为松弛的具有较低包装比环形串珠状结构。可见染色质丝在体外不同条件下可以采取一种以上的形式。除去组蛋白，可以产生负超螺旋 DNA（图2-22）。由于核小体的组蛋白被除去时，DNA 能够生成负超螺旋，那么当 DNA 处在核小体表面时，必然处于松弛状态。正是由于组蛋白的作用，才解除了松缠状态造成的张力。实验表明，核小体上 DNA 的结构周期 10.3 bp/圈，而不是像在溶液中那样为 10.5 bp/圈。然而核小体上 DNA 结构周期并不均匀。在中心节段的 DNA 可能高达 10.7 bp/圈。

## 五、着丝点 (centromere)

着丝点使染色体在有丝分裂时表现出形态上的不连续性，因此人们曾一度认为着丝点是两个构成染色体的 DNA 分子的连接区域。然而，染色体是一条完整的 DNA 分子的发现，人们才知道，着丝点序列乃是染色体 DNA 上的一段特殊序列，能够促进与纺锤体的相互作用。通过染色体步行法（见第六节）人们找到酵母第 3 号染色体的着丝点 CEN 3 的核苷酸序列。将 CEN 3 插入到酵母质粒中，能使质粒在细胞有丝分裂和减数分裂期间表现出有规则的向子细胞中分配的行为。所有的着丝点序列都含有大约 130 bp 长的序列，其中部序列富含 AT，两端则为两段高度保守的序列（图2-23）。在着丝点序列附近，还有高度重复的卫星 DNA。这些卫星 DNA 不能与组蛋白结合，因此形成异染色质区域。在 CEN序列所在的区域，大约 220 到 250 bp，比其他部分的染色质更能抵抗核酸酶的水解。这部分与多种蛋白质结合成为动原体(kinetochore)，动原体参与了和纺

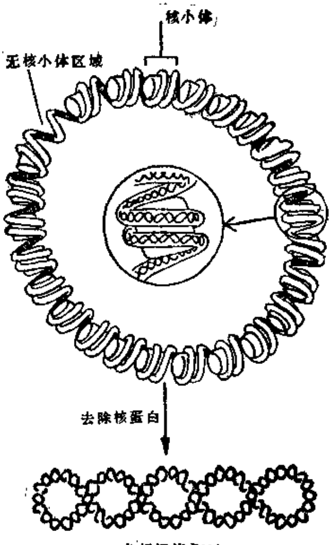



图 2-22 从SV40 的微染色体可以得到负超螺旋 DNA

| | 成分 I | 成分 II | 成分 III |
|---|---|---|---|
| CEN 3 | ATAAGTCACATGAT | ←88 bp(93% AT)→ | TGATTTCCGAA |
| CEN11 | ATAAGTCACATGAT | ←89 bp(94% AT)→ | TGATTTCCGAA |
| CEN 4 | AAAGGTCACATGCT | ←82 bp(93% AT)→ | TGATTACCGAA |

图 2-23 酵母的三个着丝点序列

锤体的微管蛋白的结合。可能动基体蛋白与 DNA 的结合妨碍 DNA 与组蛋白的结合。

### 六、端粒(telomere)

人们很早就发现，用 X 射线照射真核生物细胞时，能使染色体断裂；而断裂产生的染色体片断可以相互连接，但是天然染色体末端则不能与其他片断发生连接。这就说明天然的染色体末端结构与内部不同，即有某种结构将染色体末端封住，使之不能与其他断裂片断相连接。这种结构，人们称之为端粒(telomere)。在低等真核生物的营养核(大核)中有大量的 rDNA 分子，很容易发现其端粒结构。如尖毛虫*oxytricha*这种原生动物的营养核中就有≈25 000 个rDNA，大小从 400 bp 直到 20 000 bp。尽管这些原生动物在进化上表现出多样性，但在端粒结构上却非常相似：

(1) 都有许多短的正向重复序列，这些重复序列可以写成通用形式，$\begin{matrix} Cn(A/T)m \\ Gn(T/A)m \end{matrix}$，其中 $n=1\sim8$，$m=1\sim4$。端粒的 5′ 末端总是在富含 C 的链上，端粒的 3′ 末端总是在富含 G 的链上。

(2) 具有一条单链末端，而且这条单链总是富含 G 的链即带有 3′-OH 的单链末端。

(3) 端粒 DNA 的末端不能被外切核酸酶及单链特异性的内切核酸酶所识别，表明端粒部分的 DNA 可能受到蛋白质因子的保护。

1986年，Gottschling 和Zakian从尖毛虫的营养核中分离出两种端粒蛋白，一种为55Kd，另一种为 26Kd，它们与端粒牢固地结合，当用高浓度盐溶液将所有的组蛋白从染色质上解析出来，而端粒蛋白仍然与端粒 DNA 结合着。这两种蛋白质保护了大约 100 bp 的双链末端和 16bp 的单链尾巴。这种结合并不是共价结合，仍然依靠氢键及 van der waal 力与 DNA 结合。这种端粒蛋白(主要是 55Kd 蛋白)的结合，既需要端粒双链部分的正向重复序列，又需要 3′末端的单链序列，否则结合能力急剧下降。如果以天然的尖毛虫端粒 DNA 的结合能力为 1，则天然双链加随机单链尾巴这样的 DNA 结合能力仅为 0.02；随机双链加天然单链尾巴的DNA为 0.4；富含 G 的链(天然序列，但仅为单链)结合能力为 0.2。这些端粒蛋白，可能不仅仅是简单的结合蛋白，很可能与真核生物染色体线形 DNA 末端的复制有关。

端粒，着丝点(centromere)和 DNA 复制原点是构成染色体不可缺少的三个要素。现在人们已经成功地将尖毛虫，四膜虫的端粒与酵母的部分染色体(包括着丝点和复制原点)拼接起来再导入酵母细胞，成为人造的染色体。当这种人造染色体 DNA 大于 300 Kb 时，能正常地精确地在子代细胞之间进行分配，就像天然的染色体一样；而较小的人造染色体则常常不能精确地分配。这可能是因为较小的染色体复制太快，子代染色体相互分离过早，而不能在纺锤体上正确地排列。

酵母端粒的 DNA 重复单位为 $\begin{matrix} 5'C_{1-3}A\cdots \\ \cdots G_{1-3}T\cdots \end{matrix}$，四膜虫为 $\begin{matrix} 5'C_4A_2\cdots \\ \cdots G_4T_2\cdots \end{matrix}$，尖毛虫为 $\begin{matrix} 5'C_4A_4\cdots \\ \cdots G_4T_4\cdots \end{matrix}$。尖毛虫端粒的重复序列为：$\begin{matrix} 5'C_4A_4C_4A_4C_4TCGAGG\cdots \\ 3'G_4T_4G_4T_4G_4T_4G_4T_4G_4AGCTCC\cdots \end{matrix}$。

# 第五节　真核生物的基因[28~51]

## 一、不连续基因(interrupted 或 discontinuous genes)

自从 1977 年 Broker 和 Sharp 等人发现了不连续基因以来，现已知道绝大多数真核生物基因都是不连续基因。所谓不连续基因就是基因的编码序列在 DNA分子上是不连续的，为不编码的序列所隔开。这是人们关于不连续基因特征的初步认识。编码的序列称为外元(exon)，不编码的序列称为内元(intron)，又称为介入序列(intervening sequence)。如果一个基因有 n 个介入序列，一般总是把基因的编码部分分割成 n+1 个序列，故有内和外之称。介入序列的起源和它存在的生物学意义是一个极其诱人的研究课题，但是目前还不完全清楚，有多种假说，相互之间有不少抵触的地方，但无疑都是与基因的分子进化密切相关的。有关内容将在第七节中叙述。

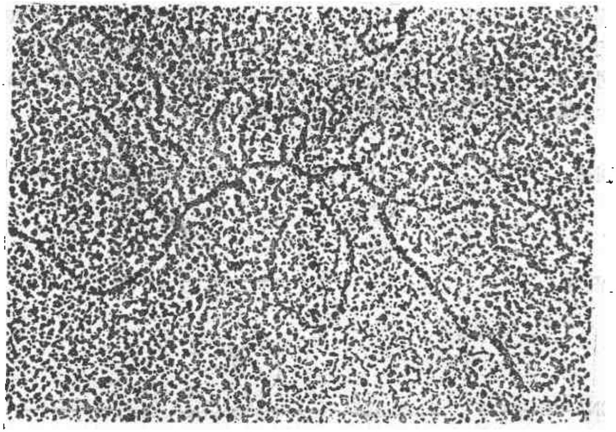
正如前言所述，"内含子"并非都"含而不显"。现在已经知道，有些内元可以编码蛋白质，而这些蛋白质的功能又是与内元序列的删除或传播扩散密切相关。早在 1980 年，Church 等人就发现了酵母线粒体细胞色素氧化酶基因的内元产物是该基因 mRNA 前体进行拼接的反式作用因子，这种因子后来称为 mRNA 成熟酶(maturase)。其细节将在转录后处理一节叙述。1985年，Michel 等人又发现酵母线粒体第Ⅱ类内元编码的蛋白质与还原病毒的逆转录酶具有很大序列相关性，这就暗示了该蛋白质在内元传播方面可能具有重要意义。同年，Jacquier 等人又推测酵母 21 S rRNA 基因的内元 r1(第Ⅰ类内元)编码的蛋白质可能具有转座酶的功能，该酶可能参与内元在基因特定位点上的整合。后来他们证明了这个 235 个氨基酸组成的转座酶确实能将这个内元转移到不含这个内元的相同基因 ($omega^-$) 中去，而对于已含有这个内元的相同基因 ($omega^+$) 则不能再发生转移。这是一个基因转换过程 (gene conversion)，其具体机制还有待于对该转座酶的生化性质作进一步的检验。1986 年，他们将该转座酶基因转换成大肠杆菌的优选密码子，克隆到大肠杆菌中去，结果大量表达了该转座酶，从而进行体内和体外的生化分析。结果表明该酶具有双链 DNA 内切酶性质，其识别位点至少包括 GGATAACA 这样八核苷酸序列。在转移过程中，似乎不需要其他的酶或蛋白质因子。

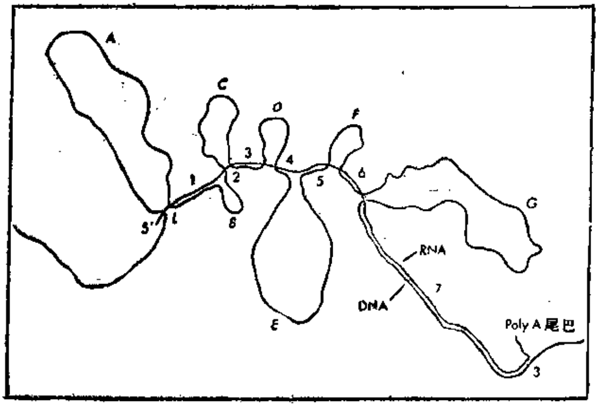
真核生物的"外显子"也并非都"显"(编码氨基酸)。除了 tRNA 基因和 rRNA 基因的外元是理所当然地不显以外，几乎全部的蛋白质基因的首尾两个外元都只有部分核苷酸序列编码氨基酸，还有完全不编码氨基酸的外元(如人类尿激酶基因的第十个外元的88个核苷酸序列)。近年来发现了通过反式拼接而外加上去的锥虫的小外元(miniexon)也完全不编码氨基酸(详见转录后处理)。

在发现内元后的相当长的一段时间内，人们曾把内元作为真核生物的"真质标记"(hallmark)。但是，Kaine(1983)和 Chu (1984)相继在原核生物中发现了内元，特别是 Chu 对于 *E.coli* 噬菌体 $T_4$ 的胸腺嘧啶核苷酸合成酶基因中 1 018 核苷酸内元的研究，从而颠覆了真质标记之说。这说明，在原核生物中原来也是存在着内元的。

由上可见，人们对于内元和外元的研究进展，已经给这两个名词充实了许多内容。值得一提的是"内含子"和"外显子"这种画蛇添足的中译名的弊病，它没有给其内涵留有进一步发展余地。故本书采用符合原文的新译名：内元和外元。

不连续基因是通过 mRNA 与 DNA 杂交试验而发现。对于杂交结果，可以用电子显微镜直接观察(图 2-24)，所形成的 RNA-DNA 杂交双链较粗，而在杂合双链的某一部位有个突出的单链环，这就是 DNA 分子上介入序列。电子显微镜只能分辨 50~100 bp，小于 50 bp 的介入序列一般不容观察到(1 μm = 3 250bp)。





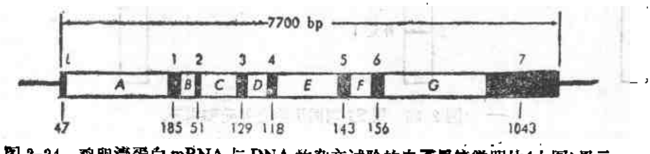




图 2-24 鸡卵清蛋白 mRNA 与 DNA 的杂交试验的电子显微镜照片(上图)显示出 8 个外元序列(L,1-7)与 mRNA 杂交，而 7 个内元(A~G)则形成嗜噗。中图为解释图。下图为鸡卵清蛋白基因中外元和内元的排列顺序及其大小(碱基对)。

另一种分析方法是所谓"Berk-Sharp 制图法"，又称"S1 核酸酶制图法"。其原理是利用S1 核酸酶的单链特异性，外切核酸酶 Ⅶ 的单链外切特异性以及稀碱溶液能分解 RNA 的特性来对杂交分子的各个部分作不同的处理以得到外元的长度，内元的长度以及基因的总长度。

图 2-25 表示将 DNA-mRNA 杂交分子作不同的处理。左边用 S1 酶处理，该酶能降解所有单链区域，留下一个带有切口的 DNA 与 mRNA 的杂交双链，如果直接走电泳可以得到一个带；相当于外元 1 加上外元 2 的长度；如果经过碱处理再走电泳，则得到两条带，一条为外元 1，一条为外元 2。右边用外切核酸酶 Ⅶ 处理，去除两端未杂交的 DNA，再用碱去除 RNA，走电泳得到一条带，为外元 1 + 内元 + 外元 2 的总长度。

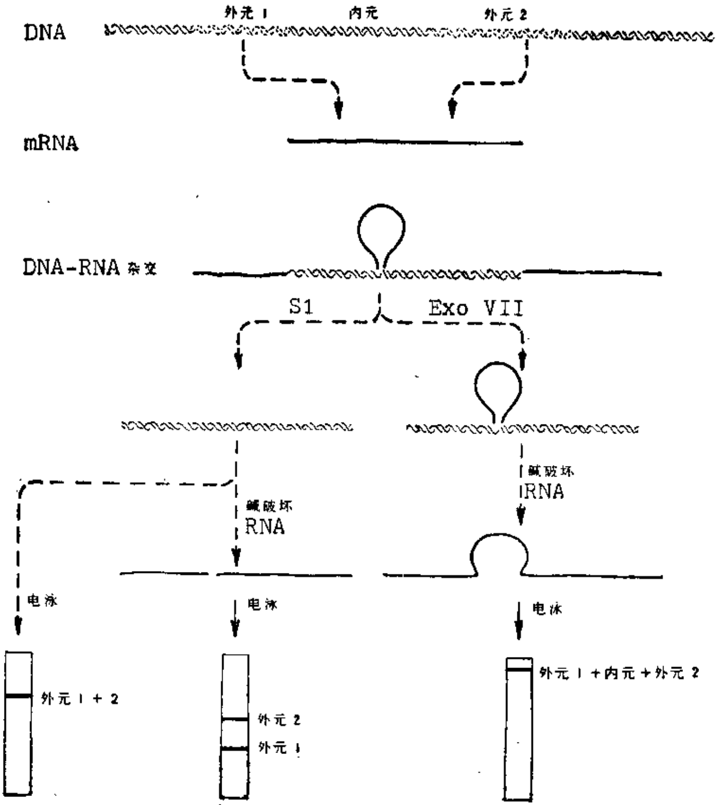



图 2-25　用 S1 制图法测定外元和内元。

不连续基因分布很广，真核生物的绝大多数结构基因都含有内元，内元是基因的组成部分，也是遗传物质，不管是什么组织，还是体细胞与生殖细胞，某种基因的内元成份都是一样的。所有的不连续基因中，外元的顺序与在 mRNA 中的顺序都是一致的，因此不连续基因仅

仅是不连续的，而不是弥散的。在核基因的内元子中，通常三个阅读框都含有无义密码子，因此一般不能编码蛋白质。除了蛋白质基因外，真核生物中编码 rRNA 和 tRNA 的基因均已发现内元的存在。在某些生物的细胞器基因中也发现内元的存在，如酵母线粒体，各种植物叶绿体。

在不同的基因中，内元的大小和数目变化很大。有的基因只含有一个或少数几个内元，如珠蛋白基因含有两个内元。对于 β-珠蛋白来说，第一个内元位于第 30 个和第 31 个密码子之间，第二个内元位于第 104 和第 105 个密码子之间（图 2-26）。不同动物的珠蛋白基因（包括 α-珠蛋白和 β-珠蛋白基因）中这两个内元位置相同，第一个内元总是很短；第二个内元要长得多，其长度在不同物种之间有所不同；而外元的长度却基本上是恒定的。珠蛋白的基因总长度的变化主要是由内元 2 的变化产生的。例如小鼠的 α-珠蛋白基因的内元 2 只有 150 bp，因而 α-珠蛋白基因的总长度只有 850 bp；而 β-珠蛋白基因的总长度为 1 382 bp。然而，它们的外元却相差不多，因此它们的 mRNA 长度变化要小得多：α-珠蛋白基因的 mRNA 长 585 b，β-珠蛋白基因的 mRNA 长 620b。

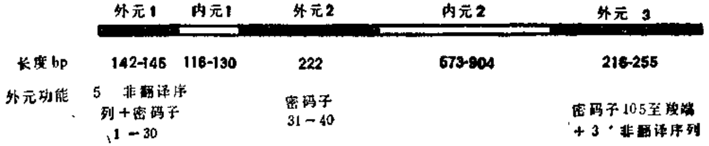



图 2-26 哺乳动物的 β-珠蛋白的基因与所有的其它活性珠蛋白基因一样，具有三个外元和两个内元。

某些基因含有较多的内元。例如鸡卵清蛋白含有 7 个内元，伴清蛋白含有 16 个内元，它们的基因总长度分别为 7 500bp 和 10 000bp，而它们的 mRNA 则只有 1859b 和 2500b。有的基因的内元特别长，例如小鼠的 DHFR 基因有 5 个内元。基因总长度为 31 000 bp，而其 mRNA 只有 2 000b。有的基因的内元和外元有更高的相嵌性，例如鸡的胶原蛋白基因有 52 个内元，53 个外元。每个外元都很短，从 45 到 249bp；；内元有的很短，比最小的外元还短，有的则很长。这样，使总长度为 5 000bp 的外元序列散布在总长度为 40 000bp 的基因序列中。这就像一连串的小岛分布在海洋中一样。这在一定程度上解释了 C 值矛盾。

在真核生物中也有些结构基因是没有内元的，如组蛋白的基因，干扰素基因等。它们大多以基因簇的形式出现。大多数的酵母蛋白质基因也没有内元。

含有内元的基因首先转录成原初转录物（primary transcript）。这样的前体 mRNA 不能翻译蛋白质；这样的前体 rRNA 或 tRNA 也不能执行其功能。在原初 RNA 转录物中相应于 DNA 内元的介入序列（有时亦称内元），必须除去以后，将外元序列连接起来，才能产生有功能的 mRNA、rRNA、tRNA，这个过程叫做 RNA 的拼接（RNA splicing）。关于内元的序列特征以及拼接机制将在第六章讲授。

## 二、基因家族与基因簇

真核生物的基因组中有许多来源相同、结构相似、功能相关的基因，这样的一组基因称为基因家族（gene family）。研究较多的基因家族是血红蛋白基因家族。

血红蛋白是脊椎动物红血球的主要成分；其功能是运送氧气和二氧化碳。每个血红蛋白分

子由两两相同的四个亚基和四个血红素组成。不含血红素的血红蛋白称为珠蛋白。所有的珠蛋白亚基的基因都有十分相似的结构。在个体发育的不同阶段，血红蛋白的亚基组成是不同的，这些不同的亚基都是由各自的基因编码的。成人的血红蛋白中，大约有97%都是$\alpha_2\beta_2$，此外还有约2%的$\alpha_2\delta_2$以及胎儿阶段遗留下来大约1%的$\alpha_2\gamma_2$；在胎儿阶段的组成为$\alpha_2\gamma_2$；在8周前的胎儿阶段，血红蛋白的组成为$\xi_2\varepsilon_2$、$\xi_2\gamma_2$、$\alpha_2\varepsilon_2$。

在以上这些组成中，$\xi$链与$\alpha$链在结构与功能上更为相似，称为类$\alpha$链；而$\varepsilon$、$\gamma$、$\delta$与$\beta$链很相似，称为类$\beta$链。在个体发育过程中，类$\alpha$链和类$\beta$链都经过一系列的替代变化，类$\alpha$链基因表达的时间顺序是$\xi\longrightarrow\alpha$；类$\beta$链基因表达的时间顺序是$\varepsilon\rightarrow\gamma\rightarrow\delta\rightarrow\beta$。这是生物体为了适应不同的环境条件而由同一祖先基因演化出来的基因家族。类$\alpha$链的基因彼此靠近成串地排列在一起，形成一个基因簇(gene cluster)，位于第11号染色体上；类$\beta$链的基因也排列成类$\beta$链基因簇，位于第16号染色体上。这是用不同发育阶段的细胞mRNA与DNA克隆作分子杂交而发现的。用这样的方法还发现了前所未知的珠蛋白的假基因(pseudogene)。许多灵长类动物的$\alpha$簇与$\beta$簇基因的组成和次序和人类几乎没有区别，都是按照"5'——胚胎血红蛋白组成亚基——胎儿血红蛋白组成亚基——成体血红蛋白组成亚基——3'"这样的次序排列的(图2-27)。

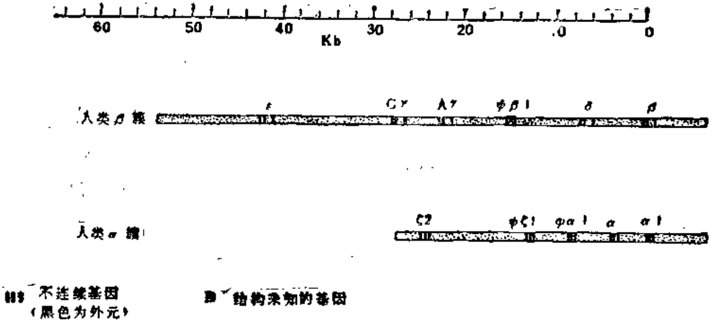



图2-27 人类类$\alpha$珠蛋白基因族和类$\beta$珠蛋白基因族各自组织成为单一的基因簇，每个基因簇都包含有功能的基因和假基因

人类的$\beta$簇分布在50kb范围的DNA上，包含五个有功能的基因($\varepsilon$，两个$\gamma$，$\delta$，$\beta$)和一个假基因。其中两个$\gamma$基因只有一个氨基酸差别，即它们的第136个氨基酸在$\gamma_G$中为Gly，而在$\gamma_A$中为Ala。$\alpha$簇比较紧凑。分布在28kb DNA的范围内，包含一个活跃的$\xi$基因，一个$\xi$假基因，两个活跃的$\alpha$基因和一个$\alpha$假基因。两个$\alpha$基因完全相同。这种两个或多个不在姐妹染色体同源位置而通常在同一个染色体上的相同基因称为非等位基因(nonallelic gene)。假基因是基因组中因突变而失活的基因，它和同一家族的活跃基因在结构上和DNA序列上有相似性。假基因没有蛋白产物，只能通过比较活跃基因和假基因的DNA序列才能发现。对于表达量很少的基因也常常是通过鉴定其DNA序列而发现的。有些生物的珠蛋白基因族不含假基因，如鸡的$\beta$簇只有14kb长，仅含4个类$\beta$活跃基因。不同生物的相应基因簇中基因和假基因的数目相差很大，以致于基因簇的长短也相差很大。

珠蛋白基因是否全集中在类α簇和类β簇呢?就目前所知,凡活跃的珠蛋白的基因都集中在类α簇和类β簇,但假基因已发现于基因组的其他部位。例如小鼠和猴类的α假基因都不在类α簇所在的染色体上。从这些假基因的结构来看,很可能由逆转录酶所介导的转座作用而产生的。有些基因族,它们的成员并不集中排列成为基因簇,而是散布在基因组中不同部位上,如果蝇的肌动蛋白基因族和微管蛋白基因族。

脊椎动物的肌红蛋白的亚基的结构与血红蛋白亚基的结构也很相似,它们也可能有着相同的来源。豆科植物中也发现了一种类似动物血红蛋白的分子,称为豆科血红蛋白。从蛋白质一级结构,内元数目和位置上看,这些珠蛋白类似物都起源于同一个祖先基因。这就表明在动植物分化之前,这种祖先基因就已存在了,而且是不连续基因。

## 三、串联重复基因

上而我们谈到了结构和功能相似,来自同一祖先基因的基因家族。他们在基因组中大多以基因簇的形式出现,也有不形成基因簇的。从广义上说,一个基因家族的各个基因也可以看成是重复基因,但其序列差异毕竟还比较大。下面我介绍真正的串联重复基因。它们的特点是:(1)各成员之间有高度的序列一致性甚至完全相同,不像基因家族中各成员差异较大;(2)拷贝数高,常有几十个甚至几百个;(3)非转录的间隔区短而一致,不像基因家族各成员之间的非转录间隔区长短不一。组蛋白基因,rRNA 基因和 tRNA 基因都是串联重复基因,这些基因的产物在细胞中都是大量需要的。

**1. 组蛋白基因** 在许多生物中,编码 H1、H2A、H2B、H3、H4 这五种组蛋白的基因彼此靠近构成一个重复单位。许多这样的重复单位串联一起,构成组蛋白基因簇。组蛋白基因簇中的重复单位数依生物而异,鸡约有 10 个,哺乳动物有20个,非洲爪蟾约有40个,果蝇约有100个,各种海胆约有 300~600 个。就目前所知,这些组蛋白基因都没有内元。组蛋白基因重复单位的组织形式也因生物而异,这表现在基因次序、转录方向和基因间隔区的不同(图 2-28)。

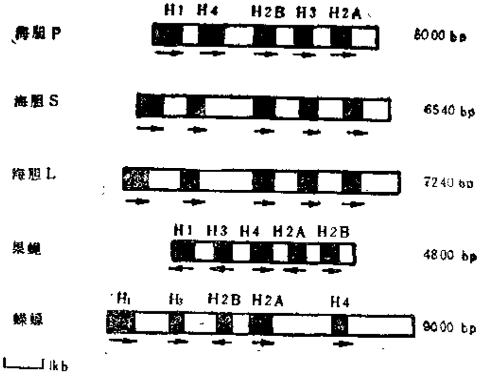



图 2-28 各种生物的组蛋白基因簇的重复单位在组织方式上有所不同
(■ 基因 □ 不转录的间隔区 → 转录方向)

海胆的组蛋白基因簇是人们发现的第一个组蛋白基因簇。从图2-28可以看出，三种不同种的海胆，其组蛋白基因簇的组织情况及转录方向完全相同，基因也基本相同，差别只在于不转录的间隔区。他们的间隔区不但长短不同，而且序列也完全不同。这就提示我们，这类组蛋白基因簇在三种海胆分化以前就已经存在。由于基因受到自然选择的压力而保持相对稳定，而间隔区没有受到这种压力因而变化较快。这种情况与不连续基因中的外元和内元的情况是类似的。实际上，在同一种海胆中，不同的重复单位中的各间隔区在大小和序列上都有轻微变化，这种变化称为轻微不均匀性(microheterogeneity)。海胆的晚期组蛋白基因大约有10个拷贝，没有固定的组织方式，而且不与早期基因簇混在一起。总的倾向是，同一物种内早期组蛋白基因成簇排列，而晚期基因弥散分布，这就表明，晚期基因可能是早期基因的某些成员通过转位而产生的。由于开始不受选择压力而发生的序列歧化，从而获得对机体有利的新功能，成为现在的晚期基因。

各种生物的组蛋白基因的组织情况各不相同。果蝇的组蛋白基因全部呈串联重复的基因簇形式，但各重复单位内基因的次序和转录方向不同。蝾螈只有一个单一类型的重复单位，但各重复单位并不靠近，而间隔着 10～50 kb 的高度重复序列。鸡的组蛋白基因重复单位内各基因无一定次序，而且重复单位并不串联在一起；哺乳动物的组蛋白基因并不都组织成为重复单位，有些散在的组蛋白基因。总的来说，组蛋白基因的组织和重复情况似乎是介于小基因簇(如珠蛋白基因簇)和典型的串联重复基因簇之间。

2. rRNA 基因　无论是原核生物，还是真核生物，它们的基因组中都含有大量的 rRNA 基因，这是与蛋白质合成中需要大量的核糖体相适应的。原核生物有 5s、16s、23s 三种 rRNA，这已讲过，真核生物中有 5.8S、18S、28S 和 5S 四种 rRNA。前三种 rRNA 称为主体 rRNA。主体 rRNA 基因拷贝数随生物而异(表 2-5)。高等真核生物一般为 200 个～500 个拷贝，三种

表 2-5　各种生物含有多拷贝的 rRNA 和 tRNA 基因

| 物　种 | rDNA 比总 DNA | 基因数 | | |
|---|---|---|---|---|
| | | 主体 RNA | 5SRNA | tRNA |
| 大肠杆菌 | 1.0% | 7 | (7) | 60 |
| 啤酒酵母 | 5.5% | 140 | (140) | 250 |
| 黑腹果蝇 | 1.3% | 250(雌)<br>150(雄) | 165 | 850 |
| 人　类 | 0.4% | 280 | 2 000 | 1300 |
| 非洲爪蟾 | 0.18% | 450 | 24 000 | 1150 |

主体 rRNA 基因组成重复单位，转录出一个 45S rRNA，然后通过转录后处理，切除转录的间隔区，成为 18S rRNA，5.8S rRNA 和 28S rRNA。一个重复单位的分子量大约为 $8\times10^6$(图

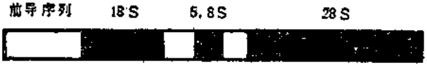



图 2-29　主体 rRNA 基因重复单位结构以及 45S rRNA 的结构(前者为 DNA 双链，后者为单链 RNA)

(2-29)。在原核生物中相应于 5.8 S 处常为若干个 tRNA 基因。

各重复单位之间的非转录间隔区在各物种之间差异很大,啤酒酵母非转录间隔区1 750bp,而小鼠则高达 30 000 bp(表 2-6)。

**表 2-6 rRNA 基因簇串联重复单位的长度(bp)**

| 物 种 | 重复单位长 | 非转录的间隔区长 | 转录区长 |
|---|---|---|---|
| 啤酒酵母 | 8 950 | 1 750 | 7 200 |
| 黑腹果蝇 | 11 500—14 200 | 3 750—6 450 | 7 750 |
| 非洲爪蟾 | 10 500—13 500 | 2 300—5 300 | 7 875 |
| 小 鼠 | 44 000 | 30 000 | 13 400 |

在细胞核中,rRNA 基因的大量转录形成了大量的 RNA,由于其染色性质与 DNA 不同,在光学显微镜下形成特殊的区域,这一结构称为核仁。含有 rDNA 串联重复单位的染色体部分称为核仁组织者 (nucleolar organizers),有多少个 rDNA 串联重复区就有多少个核仁组织者。

有些 rRNA 基因以独立的染色体外分子(extrachromosomal molecules) 形式存在,这些染色体外分子仍然处在核内。其中有两种情况:一种是长期的稳定的染色体外分子,发现于一种果蝇(D.discoideum)中(图 2-30)。这种分子是一个很大的回文序列,中间间隔20 000个bp,每一个重复单位为两个转录元,一个转录主体 rRNA 基因,一个转录 5S rRNA 基因。关于这个分子的来历还不清楚。另一种是出现在某些动物的胚胎发育过程中的暂时的染色体外分子,如在非洲爪蟾卵母细胞中的大量 rDNA 染色体外分子。这种现象称为基因扩增,其机制将在复制一章阐述。

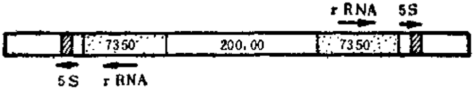



图 2-30 果蝇 D.discoideum 的染色体外分子,箭头表示转录方向

在绝大多数真核生物里,5 S RNA 基因的组织形式与主体 rRNA 基因的组织形式不同。5 S RNA 基因拷贝数很多,而且它们不是由 RNA 聚合酶 I 转录的,而是 RNA 聚合酶Ⅲ转录的,因为 5 S RNA 基因的启动子全部位于基因内。5S RNA 基因没有基因扩增现象。我们还不知道,5 S RNA 基因的表达是否与主体 rRNA 基因的表达相协调以及协调的机制。

目前研究的较多是非洲爪蟾的 5 S RNA 基因。非洲爪蟾有两类 5S RNA,一类存在于卵母细胞中,另一类存在于体细胞中。两种 RNA有 6 个碱基差异。爪蟾约有 20 000 个5S RNA 基因,其中绝大部分是卵母细胞型的 5 S RNA 基因,这可能是为了与主体 rRNA 基因扩增机制相适应的。卵母细胞型 5S RNA的基因簇的串联重复单位如图 2-31 所示。在图2-31中,左边是一个富含 A·T 序列的不转录间隔区,该区平均长约 400bp,基本上是由 15bp 的重复序列 (CAAAGTTTGAGTTTT) 及其相关序列构成的;向右是 49 bp 的富含 G·C 序列的不转录间隔区;然后就是 120 bp 的 5 SRNA 基因本身;然后是 $1^1/_2$ 个富含 A·T 序列的重复单位,接着

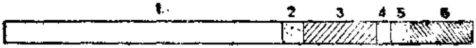



图 2-31　非洲爪蟾卵母细胞型 5 S RNA 的基因簇串联重复单位，其中
1 和 4 为富含 A·T 的不转录间隔区，主要重复单元 CAAAGTTTGAGTTTT；
2 和 5 为富含G·C的不转录间隔区；3 为 5S RNA基因；6 为 5S RNA假基因。

又是 49bp 的富含 G·C 的间隔区；最后是 101bp 的 5SRNA 假基因，它有 9 个点突变（缺失突变），这可能是造成它不能表达的原因。非洲爪蟾还有少量的另一种卵母细胞型的 5SRNA，称为微量 5 SRNA，其基因也是成为串联重复的基因簇形式，但其重复单位只有 350 bp，其不转录的间隔区与主要的 5 S RNA 基因重复单位的间隔区无相关性。

3．tRNA 基因　tRNA 约长 70～80 b，而其基因约长 140 bp。目前研究得较多的是酵母 tRNA 基因。tRNA 也是串联重复排列，但各重复单位中的各 tRNA 基因可以不相同。我们还不知道 tRNA 基因簇中各个基因单独转录还是整个重复单位一起转录。许多(并非全部)酵母 tRNA基因含有 10～20 bp 的内元，各 tRNA 基因的内元没有序列的共同性。转录后不但要切除两端的转录的间隔区，而且要进行拼接才能产生成熟的 tRNA。3′ 末端的 CCA 序列并不存在于原始转录物中，而需要外加一个 CCA 序列。拼接机制将在转录一章讲授。在进化上相距很远的有机体中，tRNA 处理酶系可能是相同的。将含有酵母 tRNA 基因的 DNA 片断注入爪蟾的卵中，结果这个 DNA 不仅被转录，而且还产生了酵母 tRNA。由于酵母和两栖类在几亿年前就从进化的道路上歧化出来，而处理酶系保持长期不变，这说明转录后处理过程在真核生物的基因表达中具有关键性的作用。

## 四、细胞器基因

在真核生物细胞器中，只有两类细胞器能够携带遗传物质，一种是线粒体，一种是叶绿体。大多数动物细胞只有线粒体，植物细胞既有线粒体，又有叶绿体。

人们早就发现了植物中的非孟德尔遗传现象，这种现象首先表现在母性遗传 (maternal inheritance)，即某些基因型只有母亲才能提供。后来人们知道，在动物的生殖细胞中，只有卵子才含有线粒体基因，而精子完全缺乏线粒体基因。还有一种非孟德尔遗传现象，叫做体细胞中的分离现象，(somatic segregation)，即在某些植物中，杂合子的两个亲本的基因型都活存下来，但在分化成体细胞时，某一部分体细胞中某种亲本的基因占优势。这就表明细胞器的基因是多拷贝的。细胞器的DNA是核外DNA，其遗传物质仍然是 DNA。过去常叫做细胞质遗传(cytoplasmic inheritance)是不恰当的。尽管细胞器位于细胞质中，然而细胞器是一个相对独立的单位，其蛋白质合成均在细胞器内进行，再说，其遗传物质并不是蛋白质。在真核生物的线粒体和叶绿体中，除纤毛虫线粒体 DNA 为线状以外，其余均为环状的非重复 DNA 序列。一个细胞器中一般含有几个或几十个相同的 DNA 分子。每个细胞中常含有许多线粒体或叶绿体，故细胞中细胞器 DNA 数量可能非常之大(表 2-7)。叶绿体 DNA 一般较大，在被子植物中约有 135～160kb，老鹳草的叶绿体 DNA 是目前已知最大的一种(217kb)。低等植物中的叶绿体 DNA 也常常高达 200kb。哺乳动物的粒线体约 16.5kb，果蝇和爪蟾约 18.4kb，酵母线粒体 DNA 高达 84kb。植物线粒体大小差异很大，有的证据表明植物线粒体中含有一种以上的 DNA 分子。

表 2-7 真核生物细胞器 DNA(叶绿体的复杂长度变化小，而线粒体的复杂长度变化大)

| 细胞器 | 物种 | DNA 复杂长度 (kb) | 每个细胞器中基因组数目 | 每个细胞中细胞器数目 | 细胞器 DNA 占总 DNA 的百分数 |
|---|---|---|---|---|---|
| 叶绿体 | C. reinhardii | 195 | 70～100 | 1 | 14% |
| | E. gracilis | 135 | 40 | 15 | 3% |
| | 高等植物 | 140 | 不详 | 不详 | 不详 |
| 线粒体 | 小鼠 L 细胞 | 16.2 | 2 | 500 | 0.2% |
| | 人 HeLa 细胞 | 16.6 | 10 | 800 | 1% |
| | 黑腹果蝇 | 18.4 | 不详 | 不详 | 不详 |
| | 非洲爪蟾 | 18.4 | 不详 | 不详 | 不详 |
| | 啤酒酵母 | 84 | 4 | 22 | 18% |

线粒体和叶绿体均编码自身所需的某些蛋白质以及 tRNA 和 rRNA，其余所需的蛋白质及酶均由核基因编码，在细胞质中合成好以后输入线粒体和叶绿体。线粒体膜和叶绿体膜都不允许核酸分子通过，因此二者基因的表达所需的 RNA 均由其本身提供。细胞器的 mRNA 只能在细胞器内翻译，它不能透过细胞器膜进入细胞质。然而近来人们发现哺乳动物线粒体中有一种类似于 RNAaseP 的 RNA 转录后处理酶中的 RNA 成分，是由核基因编码的。这种 RNA 分子是怎样能够穿越线粒体膜的，现在还不清楚。

细胞器有自己的蛋白质合成机器。上文已经说到，rRNA 和 tRNA 均由细胞器自己提供。rRNA 只有两种，各自与蛋白质构成大亚基或小亚基；全套的 tRNA 均比核基因编码的 tRNA 小。氨酰-tRNA 合成酶，RNA 聚合酶和核糖体蛋白质均由核基因编码，然而它们都不同于细胞质中蛋白质合成机器所用的组分，是为细胞器专用的。酵母线粒体蛋白质合成机器所需的组分列于表 2-8，其他生物的线粒体和叶绿体的蛋白质合成机器都与此相似。

表 2-8 酵母线粒体蛋白质合成机器的装配

| 组分 | 线粒体内合成的 RNA | 细胞质内合成的蛋白质 |
|---|---|---|
| 模板 | ～8 种 mRNA | RNA 聚合酶 |
| 核糖体 | 2 种 rRNA | ～75 种核糖体蛋白质 |
| 应接体 | 22 种 tRNA | 22 种氨酰-tRNA 合成酶 |

酵母线粒体 DNA 为84kb，其上面的基因有如下几个特点：①存在大量的(占基因组25%)由短序列组成的富含 A·T 的不编码区域；② 在其中两个 rRNA 基因中，21 S rRNA基因含有一个内元，而 15S rRNA基因没有内元；③ 有两个重要的基因是不连续基因，一个是细胞色素 b 基因，一个是细胞色素氧化酶亚基 1 基因。其中有不少长的内元具有与前一个外元相一致的开放阅读框(open reading frame)，在某些情况下，这些内元能够表达成蛋白质，尽管数量较少，但对该内元的删除以及内元的散布又是必不可少的(详见第六章)。

哺乳动物的线粒体与酵母的线粒体有很大的不同，长度只有 16 kb 左右，基因安排得十分

紧凑，没有内元，有些基因重叠，几乎每个碱基对都被分配到基因中去。人类的线粒体 DNA 序列已经清楚，共有 16 569 bp，含有 37 个基因(图 2-32)。下面，我们从基因的排列、转录、翻译、密码子的使用等方面简要地讲述人类线粒体的分子遗传学特点。关于线粒体的复制将在第三章中讲述。

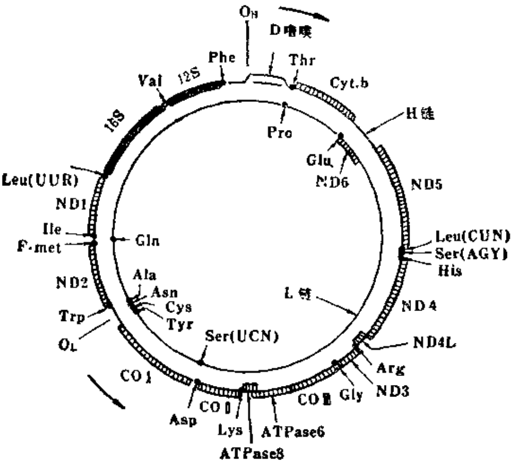



图 2-32　人类线粒体基因图谱。整个图谱分为两个部分：外圈表示从重链(H 链)上转录的基因分布情况，内圈表示从轻链(L链)上转录的基因分布情况。两个 rRNA 基因用黑色框表示，22 个 tRNA 基因用黑色圆点表示；蛋白质基因用短线框表示。ATPase 6 和 8 为线粒体 ATP 酶复合体的组成成分；CO Ⅰ，CO Ⅱ，CO Ⅲ 为细胞色素氧化酶的亚基；Cyt b 表示细胞色素 b；ND1-6 和 ND4L 表示呼吸链 NADH 脱氢酶的组分(过去称为未经鉴明的开放阅读框 URF1-6 和 URF4L)。$O_H$ 和 $O_L$ 分别表示复制时新生的 H 链和 L 链的复制起点。

(1) 基因的组织情况：有 13 个蛋白质基因，其中一个是细胞色素 b 基因 (cyt b)；两个是 ATP 酶亚基的基因(ATPase6，ATPase8)；三个是细胞色素氧化酶亚基的基因(CO Ⅰ，CO Ⅱ，CO Ⅲ)；七个是呼吸链 NADH 脱氢酶亚基的基因(ND1-6，ND4L)。有两个 rRNA 基因，一个是 12 S rRNA，一个是 16 S rRNA。还有 22 个 tRNA 基因。在重链上，靠近的蛋白质基因以及 rRNA 基因之间大多有 tRNA 基因散布，在 ATPase8—ATPase6—CO Ⅲ 之间以及 ND4L—ND4 之间没有 tRNA 基因。

人类线粒体的基因排列得非常紧密。在 37 个基因之间，除了线粒体 DNA 复制所不可缺少的 D-噜噗 区域之外，基因间隔区总共只有 87bp，只占 DNA 总长度的 0.5%。在许多情况下，基因之间没有间隔。即前一个基因的最后一个碱基与下一个基因的第一个碱基相衔接。有时有基因重叠，但大多数只重叠一个碱基，也有重叠较大的，如 ND4L 和 ATPase8 与它们相邻的基因有较大的重叠。为了有效利用 DNA，有五个阅读框缺少终止密码子，仅以 U 或

UA 结尾,因此其 Ochre 密码子要借助于 mRNA 上的 poly(A)尾巴产生出来。有三个蛋白质基因的终止密码子为 AGA 或 AGG,这在核基因密码子表中为精氨酸密码子。

(2) 转录:在 13 个蛋白质基因中,只有 ND6 从轻链上转录,其余均从重链上转录。两个 rRNA 基因亦从重链上转录。在 22 个 tRNA 基因中,有 14 个从重链上转录。这样,线粒体的 37 个基因中有 28 个从重链上转录。由于相邻的蛋白质基因或 rRNA 基因之间插入了 tRNA 基因,因此,没有启动子立足的余地。从重链上转录的全部基因构成一个转录元,从轻链上转录的全部基因构成另一个转录元。两个启动子都位于 D-噜噗内。tRNA基因插入蛋白质基因及 rRNA基因之间,其意义在于转录后处理。由于 tRNA 具有二级结构,可为特定的 RNA 内切酶所识别,因此一下子就产生了 rRNA 分子和单基因的 mRNA,只有一个双基因 mRNA 和一个三基因 mRNA。看来,转录后处理是线粒体基因表达的中心环节。rRNA要比 mRNA 合成得多,这可能是在两个 rRNA 基因转录后由于某种蛋白质因子在特定的序列上造成转录的过早中止。轻链的转录及转录后处理情况与重链相类似。常可以在线粒体中分离出很大的轻链转录物,而得不到大的重链转录物。这暗示着在转录后重链转录物立即在 tRNA 连接处得到处理,而轻链转录物由于某种机制而没有及时处理。

(3) 翻译:通过转录后处理产生了 mRNA,由线粒体自身的蛋白质合成机器进行翻译。翻译的起始密码子可以是AUG,AUA 或 AUU,全都在离 mRNA 5′ 端 6 个碱基的距离内。加上 mRNA多以终止密码子或不完全的终止密码子结尾,因此人类线粒体蛋白质基因上的不翻译序列极少。

从以上各特点,特别是基因组织方式来看,人类线粒体非常像原核生物的噬菌体。能抑制原核生物蛋白质合成的抗生素,如红霉素、链霉素、氯霉素等均可抑制细胞器的蛋白质合成;而能抑制真核生物蛋白质合成的物质如亚胺环己酮 (cycloheximide) 对细胞器的蛋白质合成则无作用。所有这些迹象都表明,细胞器很可能是由原核生物侵入了真核生物的祖先(即原始真核生物,urearyotes),形成了极其稳定的共生关系,进而成为真核生物不可缺少的组成部分。

## 第六节 基 因 定 位[2,52~57]

基因在某一染色体上的排列叫做遗传图,确定某一基因在染色体上的位置就叫做基因的定位。这里介绍三种基因定位方法,一种是遗传交换定位法,一种是接合定位法,另一种是 80 年代才发展起来的染色体步行法和染色体跳跃法。前两种是经典的遗传学定位方法,最后一种才是分子遗传学的基因定位方法。终极的基因定位方法乃是 DNA 的序列测定。

### 一、遗传交换定位法

遗传交换发生在减数分裂时期,两个同源染色体特异地吸引而配对,然后在同一点断裂,然后交叉、复合,就形成了既有来自父本的基因又有来自母本的基因的重组染色体。染色体交换的具体机制仍然是一个谜,有多种假说,这将在重组一章介绍。

用遗传交换给基因定位的原理是:当两个基因位于不同的染色体上时,它们在单倍体分离子中的分配是随机的,也就是说它们保留在同一个分离子中的机会是50%。相反,如果两个基因位于同一个染色体上,则倾向于分配在一起,除非通过交换而分开。由于交换是沿着整个染色体纵向随机发生的,因此,相隔较远的两个基因之间断裂而引起遗传重组的机率就比较大。

也就是说，重组体出现频率越大，基因相距越远。

例如同一个染色体上的三个基因 a、b、c 可以通过双因子杂交而确定基因排列。例如：AB 和 ab 杂交，可以得到两种亲本基因型（AB，ab）和两种重组体基因型（Ab，aB）。如果这种重组体出现频率是 30%；AC 和 ac 杂交时重组体出现频率是 10%；BC 和 bc 杂交时重组体出现频率是 25%，则基因排列如图 2-33 所示。

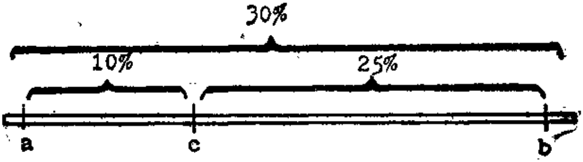



图 2-33　根据三次双因子交换推测 a、b、c 三个基因的排列顺序

用三因子杂交可以证实上述双因子杂交所提供的基因顺序的正确性。现将ACB和acb 杂交的三类重组体的基因型和出现频率示于图 2-34。由于存在着多交换，上例 a 和 b 之间重组频率 $f_{ab}$ 总是小于 a、c之重组频率 $f_{ac}$ 与 c、b 之重组频率 $f_{cb}$ 之和，即 $f_{ab}<f_{ac}+f_{cb}$。在 c、b 已经发生交换的情况下，a、c 再发生交换的机率为 $f_{cb}\times f_{ac}$；相反，在 a、c 已经发生交换的情况下，c、b 再发生交换的机率为 $f_{ac}\times f_{cb}$，因此，$f_{ab}=f_{ac}+f_{cb}-2f_{ac}\times f_{cb}$。然而在实际标定基因位置时往往受到一些因素的干扰，使得互相关联的交换频率时高时低。

在遗传图上常用图距单位来表示。如果两个基因之间重组频率是 5%，则两个基因相隔 5 个图距单位。

## 二、接合定位法

细菌的性别是由致育因子（F）的有无决定的，有 F 因子的细胞是雄性（$F^+$ 细胞），没有 F 因子的细胞是雌性（$F^-$ 细胞）。F 因子能像温和噬菌体一样，以两种状态存在于细菌中，一种是游离状态，一种是整合状态。当 F 因子处于游离状态时，细胞只是一个潜在的雄性细胞，不能和雌性细胞交配（接合），能够转移的只是 F 因子本身。只有当游离的环状性因子 DNA 整合到细胞染色体上时才能与雌性细胞交配。整合有F因子的细胞就叫做 Hfr（高频重组）细胞。由于 F 因子整合的细菌染色体部位不同，因而有各种不同的 Hfr 株系。接合过程见图 2-35 所示。

已整合有 F 因子的雄性细胞染色体，由 F 因子所编码的酶特异性识别 F 因子上的 *ori* 序列并切断一条单链，该酶并结合于 5′ 末端。这一单链末端首先进入

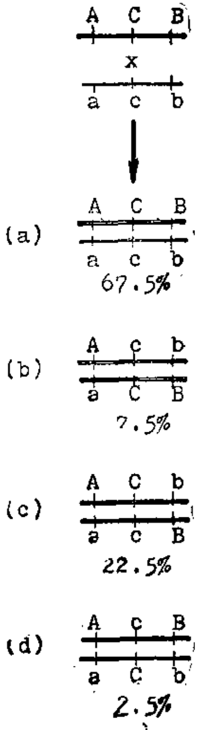



图 2-34 用三因子杂交确定基因顺序，各类百分比系图2-33中例子的理论数值
(a) 亲本基因型，(b) a,c 之间单交换；(c) c,b 之间单交换；(d) a,c之间和c,b 之间双交换。

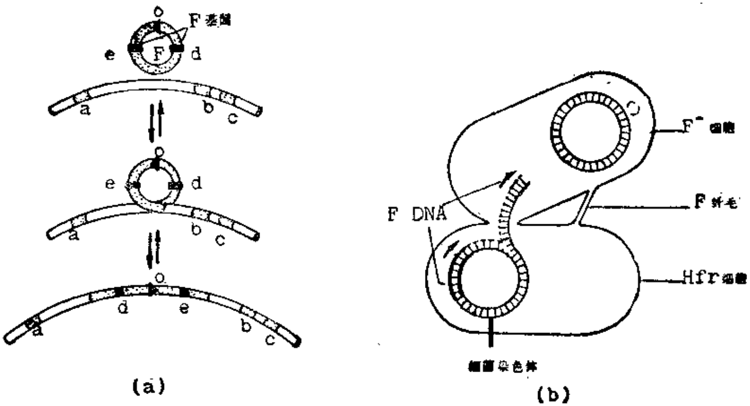



图 2-35　大肠杆菌的性过程

(a) F因子与染色体的整合　(b) Hfr细胞的染色体在一段FDNA的引导下进入$F^-$细胞。图中FDNA的比例扩大了许多倍，实际上只占Hfr细胞染色体的很小的比例。

雌性细胞。人们可以用特殊的搅拌器（如 Waring Blendor）来控制接合时间。在短时间就能进入雌性细胞的基因距离端点就近，相反，需要长时间才能进入雌性细胞的基因距离端点就远。实验证明，在37℃时，转移一条完整的大肠杆菌染色体所需的时间随菌株而异，从89分钟到100分钟不等，但对同一种菌株来说，转移速度是相当稳定的。对于大肠杆菌$K_{12}$来说，恰

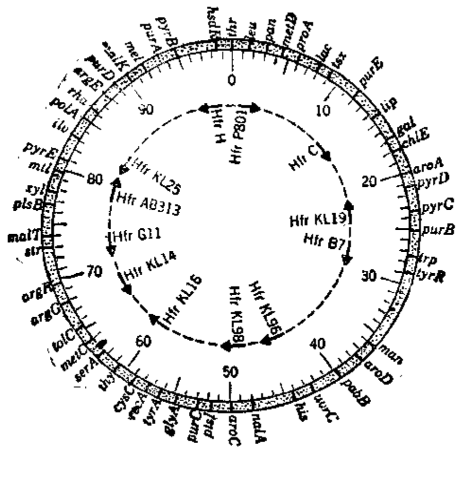



图 2-36　大肠杆菌 K12 的遗传图（外圈所示为大肠杆菌基因，内圈所示为几种了解得比较清楚的Hfr株系的F因子整合位点及转移方向）

好需要100分钟。以时钟的12/0点处作为大肠杆菌遗传图的0点，有时也写作100/0，按顺时针方向将基因定位。如果以thrA基因为0点，大肠杆菌$K_{12}$的遗传图则如图2-36所示。这样的图又称为环形连锁图。图中仅给出900多个已经定位的基因中的52个基因的位置。这52个基因定位的精确度最高，可作为进一步研究待定位基因的参照点，也可供学生初步熟悉大肠杆菌的遗传标志之用。图中用以代表基因的符号请参考B.J.Bachamann等人的评论：Bact.Rev.40:116,1976。

### 三、染色体步行和染色体跳跃

随着限制性内切酶的发现和广泛使用，染色体上的基因定位方法已发生根本性的变革，大大地提高了基因定位的速度和精确性。这种方法就是染色体步行法(chromosome walking)以及最近发展起来的染色体跳跃法，(chromosome jumping)，它是对染色体步行法的改进和补充。本书限于篇幅，只对基本原理加以介绍。

要直接分离染色体上某个基因或某一段DNA序列，现在还不可能。人们采用相反的战略：先将整个基因组DNA切成宜于操作的片断；然后与适合的载体(如质粒和噬菌体)重组，将重组载体导入寄主细胞(如*E.coli*)；让寄主细胞长成克隆。这样，一个克隆内的每个细胞的载体上都包含有特定的基因组DNA片断。这样的一套克隆就叫做基因组克隆(genomic clones)；其中克隆的一套基因组片断就叫做基因组文库(genome library)。基因组文库的大小(即所需克隆的片断数目)随载体对外源DNA的容纳能力，基因组的大小以及所要求的基因组DNA在基因文库中出现的概率而变化。一般λ噬菌体对外源DNA有较大的容纳能力。要求的概率一般为99%。这样，*E.coli*基因组需要1 500个克隆片断，酵母基因组需4 600个，果蝇需要48 000，哺乳动物需要800 000个。这样的基因文库都已经以各种方式建立起来。

基因组文库一旦建立起来，只要繁殖这些克隆，就可以通过一个或几个探针找到你所需要的克隆。这里所谓探针就是用放射性标记的能与你的目的DNA片断进行杂交的核酸分子，通常使用50bp左右的cDNA分子。筛选的方法是让不同的克隆长在培养皿里，然后再制作一个完全相同的复制皿(replica plate)。先把原皿(master plate)保存备用，然后用复制皿进行原位杂交试验(in situ hybridization)。先将带有重组载体的细菌克隆置于硝酸纤维素滤膜上溶菌，将DNA在原位变性并固定于滤膜上，然后与放射性标记的单链探针分子杂交，洗掉未杂交的探针分子，最后通过放射自显影确定探针与哪一个克隆杂交。再从原皿中挑出同样的克隆进行培养以作进一步的研究。

真核生物的基因常常很大，在制作基因组文库时常会被打断，使之分散在两个或两个以上的克隆中。除了可以应用几个不同探针寻找以外，还可以通过已经找到的克隆，取其外源DNA的末端片断进行亚克隆(subclone)，然后分离这段DNA，经过放射性标记，作为探针再去与基因组文库中的其他克隆进行杂交，以寻找与原来克隆中的外源DNA片断有重叠的基因组克隆(在制作基因组文库时就保证了各片断之间有相重叠部分)。利用这种方法不但可以重建一个大的基因，而且可以鉴定染色体上一个很大的区域。即从染色体上某一已知的起始点，可“步行”到染色体上的另一个区域并考察所经过的路程(如限制酶图谱甚至进行序列分析)。这种方法因此称做染色体步行法。简而言之，染色体步行法是通过相互重叠的基因组克隆之间进行一连串的杂交而实现的。图2-37为染色体步行法的示意图。

按道理用染色体步行法似乎可以在染色体上走完全程，但实际上是不可能的，目前最多只

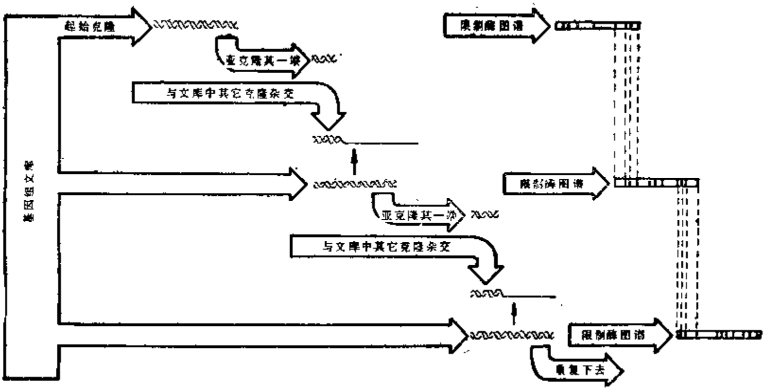



图 2-37 通过相互重叠的基因组克隆之间的连续杂交而实现染色体步行

能走 1000kb。人们所碰到的实际问题是：真核生物基因组中有许多重复序列，常使步行发生差错；大数目的基因组克隆使得每前进一步都要做大量的筛选工作，可以说是步履艰难；基因组片断在克隆过程中在大多数情况下保真度很好，但也有发生缺失情况的。

人们针对染色体步行法所用的基因组克隆数太多的缺点，采用了 C.R.Cantor 1984 年所创造的脉冲电场梯度凝胶电泳法(PFG)分离大片断 DNA(数以百万 bp 计)，设计出染色体跳跃文库 (chromosome jumping library)，常常可以跳数十万 bp。制作这种文库时，一般使用识别序列长的限制性内切酶，如 Not 1，其识别序列为 GC↓GGCCGC，若按理论计算(四种核苷酸等量)，每 $4^8$(约 66 000)bp 才有一个切点，实际上还要少得多，因为高等真核生物的基因组中 G+C 含量少于 A+T。用这样的酶分别降解基因组 DNA 和带有琥珀型抑制 tRNA 基因的质粒，使二者在适当的浓度比之下进行重组，然后连接成环，再用一种切点较多的限制性内切酶(但在质粒序列内没有切点)将大部分基因组 DNA 切除，再将质粒和其两端的少量基因组 DNA 重组到含有多点琥珀型无义突变的 λDNA 中，导入不含琥珀型抑制 tRNA 的 *E.coli* 中，这样就制成了染色体跳跃文库。制作染色体跳跃文库的步骤示于图 2-38。染色体跳跃文库的用法就不在这里介绍了，请读者自己阅读参考文献。

## 第七节 基因的分子进化[58~66]

前面我们已经谈到卫星 DNA 的进化问题。基因的进化机制也是类似的，通过基因扩增和非均等交换形成多拷贝基因，再通过突变积累、基因重排和自然选择等因素形成多成员的基因家族或形成新的基因。这些过程可以反复地进行。下面，我们主要讲述外元重排和不均等交换对基因进化的影响。

早在 1978 年，W.Gilbert 就指出，通过外元的重排(exon shufling)，能够产生新的基因，从而产生新的基因产物，这种产物具有新的功能，从而使生物体能更加适应复杂的自然环境。而内元就是这种外元重排的历史痕迹，而且内元的存在可能以某种方式促进了这种外元

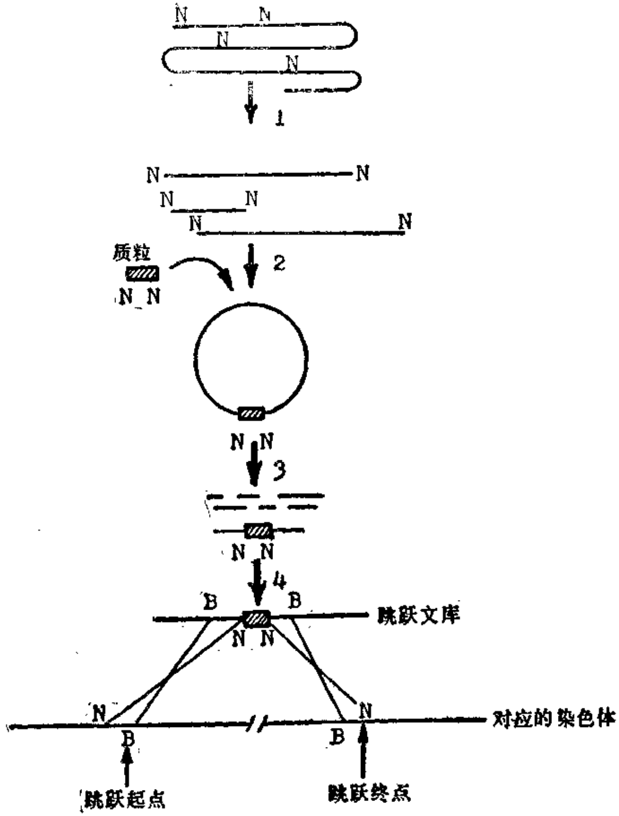


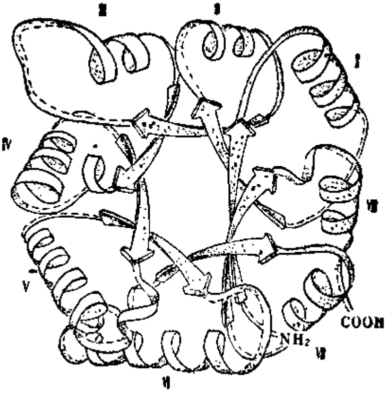

图 2-38 染色体跳跃文库的制作步骤

1. 用 Not1 切染色体 DNA，然后走琼脂糖 PFG 电泳； 2. 熔解琼脂糖，洗出DNA并与经 Not 1 酶切的质粒相连接； 3. 用识别序列较短且质粒又缺少此种位点的酶如 BamH1 进行切割； 4. 将所得片断与经 BamH1 切割的载体 λ 进行连接，这样就得到包含 Not1 位点与 BamH1 位点之间序列的跳跃文库。

重排过程。今天，大量的证据表明，在脊椎动物出现之前的 5 亿年中，外元重排在基因构建过程中起着主要的作用。这方面最显著的实例就是类表皮生长因子结构域 (epidermal-growth-factor-like domain) 在许多功能上不相关的蛋白质中的散布，这些蛋白质包括表皮生长因子前体，尿激酶原，凝血因子 Ⅸ 和 Ⅹ，低密度脂蛋白受体等。

关于内元如何起源的问题，目前主要有两种假说。以哈佛大学 W.Gilbert 为代表的一派认为，内元与它所在的基因一样古老，在装配第一个这样的基因时，内元就已存在。而今天的原核生物和少数低等的真核生物，由于它们需要进行快速的 DNA 复制从面进行快速的细胞分裂，因而失去了内元序列。而对于绝大多数真核生物来说，它们没有必要进行快速的 DNA

复制和细胞分裂，因而将内元序列保留下来，它对于保持基因的相对稳定又有重要作用(下面就要讲到)。这种观点认为，今天高等真核生物基因中的内元更具有代表性。原核生物中内元的发现有力地支持了这种假说。为了更进一步证明这种假说，人们分析了在原核生物和真核生物中广泛分布的而且序列高度保守的一个酶的基因结构，即丙糖磷酸异构酶(TIM)的基因结构。人们发现，在构巢曲霉(*Aspergillus nidulans*)中，该基因有 5 个内元，在鸡中有 6 个内元，而在玉米中有 8 个内元(图 2-39)。



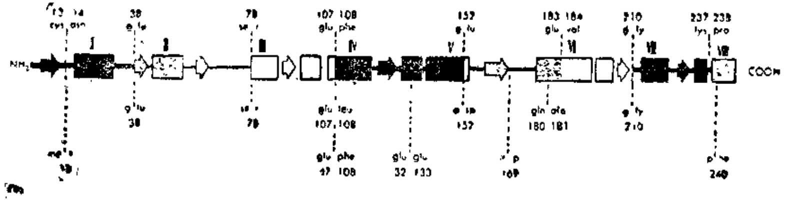




图 2-39　三种生物的丙糖磷酸异构酶(TIM)的分子结构和基因结构。
(a) 由 α-螺旋和 β-折叠所组成的三维结构；　(b) 蛋白质分子中相间的 β 折叠区(箭头)和 α-螺旋区(长方块)，竖直虚线表示内元位置(上为玉米，中为鸡，下为构巢曲霉)。

以变化缓慢的 5S rRNA基因为标准，真核生物大约在 15 亿年前从原始细胞(progenote)中分化出来；12亿年前真菌(包括酵母菌和曲霉)就与动植物祖先分家；10亿年前，动植物分家；4 亿年前出现鱼类；3 亿年前出现了哺乳类和鸟类。从图 2-39 中可以看到，在鸡和玉米的TIM基因的内元中，有 5 个内元在两种基因中的位置完全相同。很难设想，这些内元是在动植物分化以后插入的而且插入得如此精确！比较合理的解释是这些内元在动植物分化之前就已存在了，其中一个内元后来移动了三个密码子，两端的两个内元丢失了，于是出现了鸡中的情况。

以 F.Crick 为代表的一派认为，不连续基因的内元不是该基因本来就有的，而是在进化的某一过程通过转座作用插入到连续基因中去的。前面所提到的酵母线粒体基因的内元编码的转座酶所产生的基因转化，有力地支持了这种假说。1987 年，Littman 等人发现人和鼠的 T

细胞表面糖蛋白 CD4 分子中，有一段免疫球蛋白分子中的类 V 结构域(V-like domain)；而在 CD4 基因的类 V 结构域的编码序列中存在着一个内元，可是在含有类 V结构域的其他众多成员中至今尚未发现有这样的内元序列。这很有可能是在 CD4 基因进化过程中（有可能在哺乳动物分化出来以后），在 V 结构域插入此内元。

上述两种假说，虽然第一种假说获得了广泛的支持，但后一种假说也具有许多实验证据。看来要得到一个统一的正确的理论，还有待于对 RNA 拼接机制以及有可能的内元转座机制作更为详尽的研究。

人们早就注意非均等交换在基因进化中的作用。我们已经看到，各种哺乳动物中的珠蛋白基因都成簇排列，但在活跃基因和假基因的数目上有很大差异。这主要是由非均等交换而产生重复基因，其机制与重复序列的产生相同。在不连续基因中，由于外元受到选择的压力，因而突变机率很小，而内元不受这种压力，因而突变机率大。这样，交换往往在外元部位发生。由于内元在序列组成和大小方面分化很快，从而限制了这种非均等交换的发生，从而保持了基因簇的相对稳定。这可能是内元存在的生物学意义之一。然而，非均等交换是不可避免的，从人和鸡的珠蛋白基因簇的人工克隆中经常可以分离到缺失的基因簇。由不均等交换而产生的缺失和重复对机体往往是有害的，就被自然淘汰；但也有个别的不均等交换结果产生了更大的选择优势，产生了新的活跃基因，原有的基因或丢失，或保留，有的因突变而变成假基因。哺乳动物现在的珠蛋白基因簇的状况就是这种不均等交换、突变、重排和自然选择的结果。

根据各物种的珠蛋白基因组分，可以推测珠蛋白祖先基因的可能进化途径(图 2-40)。

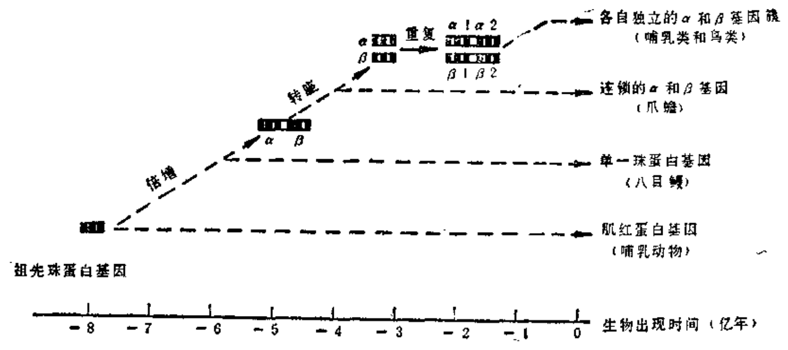



图 2-40 珠蛋白基因由一个祖先基因通过倍增、转座、重复而进化的可能途径

哺乳类和鸟类都具有各自独立的 α 基因簇和 β 基因簇；而两栖类（如爪蟾）的 α 基因和 β 基因仍然是连锁的。哺乳类和鸟类从爬行类祖先歧化出来的时间大约在 2 亿 7 千万年前，而爬行类从两栖类祖先歧化出来的时间大约在 3 亿 5 千万年前。这样，从 3 亿 5 千万年前到 2 亿 7 千万年前的时间可以定为 α 基因和 β 基因失去连锁的时间。α 基因和 β 基因分离可能是由转座作用造成的。某些原始鱼类如八目鳗只有一种珠蛋白基因，这些鱼类一定是在珠蛋白基因倍增并分化成 α 基因和 β 基因之前从进化路线上歧化出来，这个时间大约在 5 亿年前。也就是说，珠蛋白基因的倍增和分离是出现在 5 亿年前至 3 亿 5 千万年前之间。从氨基酸组成顺序的差别可以推算出珠蛋白基因和肌红蛋白基因分化的时间在 8 亿年前。豆科血红蛋白

基因与珠蛋白基因分化的时间则更早。

把不同的生物中某种蛋白质的氨基酸组成顺序(或其基因的碱基序列)进行比较，发现两种生物之间差异的百分数大约和它们自同一祖先向不同方向演化的时间成正比，也就是说，突变在生物进化过程中发生和积累的速度大体上是恒定的，这种情况常常被称为进化钟(evolutionary clock)。通过比较不同生物种类的珠蛋白基因的变化情况，就可以得出珠蛋白基因的进化速度大约为每1040万年有1%的碱基出现替代型差异。用这一速度就可以推算珠蛋白基因在种内形成变化(也可以称进化)的大致时间表。例如人类$\beta$簇中的$\beta$基因与$\delta$基因间有3.7%的碱基替代，因此这两基因分化出来的时间应在$1040 \times 3.7 = 3850$万年前。

不同的基因承受着不同的负选择压力，这就决定了生物钟的快慢，因此不可用一种基因的进化速度去推测另一种基因的进化时间表。一般来说，基因的产物越重要，对它的结构和功能要求越严格，变异积累的速度越低。同一个基因的不同区段的重要性不同，所承受的负选择压力也不同。因此外元的序列变化就慢得多。同是外元，决定蛋白质活性部位的碱基序列(从空间结构上看)比其他序列要稳定得多。

基因的重复减弱了负选择压力，使相同基因中的一个或几个能较快地积累突变从而形成新的基因家族成员或变成失活的假基因。一旦成为假基因以后就完全不受选择压力的影响，变异积累的速度就大大地高于活跃基因而最终面目全非。假基因的命运或者是因DNA缺失而消失，或者是作为一般无用的DNA单一序列而存在下去。然而，有些基因在重复后的长期进化中却一直保持着几乎完全一致，如人类珠蛋白的两个$\alpha$基因和两个$\gamma$基因。现在普遍认为，这种非等位的相同基因在遗传过程中是作为一个单位而传递的，它称为一致进化(coincidental evolution)，有时也称作共进化(coevolution)。解释一致进化现象的两种机制即交换固定学说和基因转换学说，都已在前面讲过。不管哪一种机制，其实际结果是：突变的基因要么被淘汰，要么占据全部的原有相同基因位置。

同一基因家族的不同成员之间的碱基差别是由突变(包括碱基替代突变，缺失突变，插入突变等)引起的。这些突变的频率很低，产生的后果大多有害，很快被负选择压力所淘汰。也有一部分被固定到生物群体中。那么如何解释从个体发生的突变扩展到群体或整个物种中呢？经典的理论是达尔文的适者生存理论，即突变体具有更大的正选择优势，适应力强，繁殖得更多更快。而一种由S.赖特提出的遗传漂变学说(genetic drift)认为，群体较小时，自然选择不起作用。日本群体遗传学家木村资生提出了类似的中性漂变学说。这种学说认为，从个体到群体的扩展完全是中性突变(无利亦无害的突变)在长期进化过程中随机漂变的结果。一个中性突变能否扩展到整个种群，基本上或完全取决于机会。事实上在DNA水平上的许多突变都是中性的，包括不改变氨基酸组成的同义突变以及不影响蛋白质空间结构和功能的无声突变。这一学说得到了实验群体遗传学研究上的支持。但是否真正反映生物进化的实际状况，还有待于进一步研究。

## 第八节　早期生命进化的三界系统理论[67~78]

### 一、原核生物之间的巨大差异

在本章第一节中我们叙述了原核生物和真核生物的差别以及原核生物的共同特征。正是这些资料曾使得许多生物学家对于两界系统理论深信不疑。两界系统理论不仅认为所有的生

命都可以划归这两界之一，而且认为今天的原核生物更类似于真核生物的祖先，因而真核生物就是从古代的原核生物进化而来的。在这一步一步的进化过程中，关键的步骤是演化出核膜，细胞壁的消失，细胞大小的增加以及细胞器的获得等。

然而，在原核生物各种属中，发现了许多巨大的差异。首先在生活环境上有着巨大的差异。有一种嗜热嗜酸菌(*sulfolobus*)能够生活在近于沸腾的水中(90℃)和很酸的环境中(pH2)。有一种嗜盐菌 *Halobacterium* 需要生活在近于饱和的盐水中，这些菌已发现于美国大盐湖和西南亚的死海中。其次，在代谢方面，也存在着巨大差异。大家所熟知的大肠杆菌和沙门氏菌都是通过无氧酵解或有氧的氧化磷酸化过程产生能量。在氧化磷酸化过程中，来自食物分子中的氢原子被用来还原分子氧而成为水。然而厌氧的甲烷菌则用氢来还原二氧化碳生成甲烷的过程来产生能量。硫细菌则用氢来还原元素硫成为硫化氢的过程来取得能量。绝对的厌氧菌在有空气时是不能生存的，而兼性厌氧菌在有无氧气时均能生存，如固氮菌 *Rhizobia*。但是，固氮菌只有在根瘤所提供的无氧条件下才能固氮。

除上述以外，还有许多分子水平上的差别，请见表 2-9。

**表 2-9　古细菌和真细菌之间的区别**

| 古　细　菌 | 真　细　菌 | 古　细　菌 | 真　细　菌 |
|---|---|---|---|
| 基因组经常重排<br>富含可转移成分<br>细胞壁中没有肽多糖<br>以醚键相连的分枝脂类，其中的甘油磷酸为 D 型<br>rRNA,tRNA和核糖体具有真细菌和真核生物的特征 | 基因组相当稳定<br>很少有可转移成分<br>细胞壁中有肽多糖<br>以酯键相连的不分枝脂类，其中的甘油磷酸为 L 型 | RNA 聚合酶与真核生物相似，由多亚基组成<br>嗜热菌中的"逆向"旋转酶能引入正超螺旋<br>EF2像真核生物一样对白喉毒素敏感 | RNA 聚合酶组成较为简单<br>旋转酶只引入负超螺旋<br>EF2对白喉毒素不敏感 |

## 二、基于 16 S rRNA 碱基变化的通用系统发生分类法

建立系统发生树的传统依据是细胞色素 C 的氨基酸组成的变化，因为全部的真核生物和许多原核生物都有细胞色素 C。然而，有些古怪的细菌缺乏细胞色素 C。有些细菌中虽然有细胞色素 C，但其氨基酸组成差别太大，因而无法确立它们之间可靠的亲缘关系。为了弥补这一缺陷，C.Woese 于 1969 年决定考察各种生物特别是各种古怪的细菌中的 16 S rRNA 的序列。各种生物都有 16 S rRNA；而且含量丰富，易于分析；在进化过程中高度保守，容易建立系统发生树。到了 1977 年，已有相当多的数据。于是，C.Woese 提出，原核生物应当分为两界：古细菌(archaebacteria)和真细菌 (eubacteria)，* 这两界彼此不同，就像它们不同于真核生物一样。这就是说，古细菌，真细菌和真核生物三者之间不存在谁起源于谁的问题，而是相互平行的。三者都起源于共同的祖先，这一祖先称为原始细胞(progenote)。现已发现古细菌和真细菌之间越来越多的差异(见表 2-9)。过去把古细菌和真细菌归为一类，就像把蝙蝠和鸟归

* 为了避免人们将这两界误认为都是细菌这一概念错误，C. Woese 于 1990 年提议将 eubacteria 改为 bacteriotida, archaebacteria 改为 archaeotida。

为一类一样，是一种肤浅的认识。至于原始细胞是怎样的细胞，目前并无直接的证据来描述它的性状，我们只能根据三界生物的共同特征或至少两界所具有的特征来推测原始细胞的性状。例如古细菌和真核生物都富含内元，这就表明原始细胞也富含内元，而在真细菌中由于需要进行快速的 DNA 复制以实现快速的生长而淘汰了内元。又如，真细菌和真核生物都有以酯键相连的不分枝的脂类，其中的甘油磷酸为 L 型，原始细胞也可能如此。

三界系统理论的提出，开始时遭到许多生物学家的怀疑与反对，因为这一理论与人们广为接受的但又缺乏根据的两个假设相矛盾。这两个假设是：所有的细菌都有密切的亲缘关系，细菌比任何一种真核生物都更接近于最先生存的细胞(原始细胞)。然而 Woese 却继续不断地完善他的 16 S rRNA 通用系统发生树，同时集累其他生物化学证据。今天三界系统理论已得到广泛的承认。根据 21 种生物的 16 S rRNA 序列的分析，三界系统进化树示于图 2-41。从这一系统中可以得到一些人们大吃一惊的结论，如人与玉米的关系比革兰氏阴性菌(大肠杆菌等)和革兰氏阳性菌(枯草杆菌等)的关系更为接近！目前，还没有将这种进化树与地质年代联系起来；如果要联系起来，就需要一个可能不公正的假设：所有的生物中 DNA 的突变速率都相同。

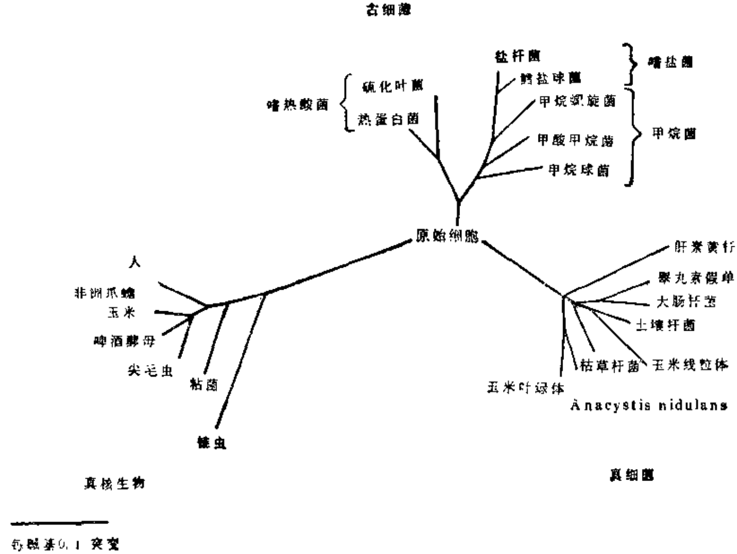



图 2-41　基于 16 S rRNA 序列分析的三界系统进化树

### 三、内共生作用参与了真核生物的进化

真核生物与原核生物的主要区别，除了核膜之外，就要算真核生物的细胞器了。核膜是如何产生的？目前还无充足的根据，大多数人认为是由细胞质膜内褶先形成内质网然后转变为核膜。至于产生这一过程的动力还毫无所知。

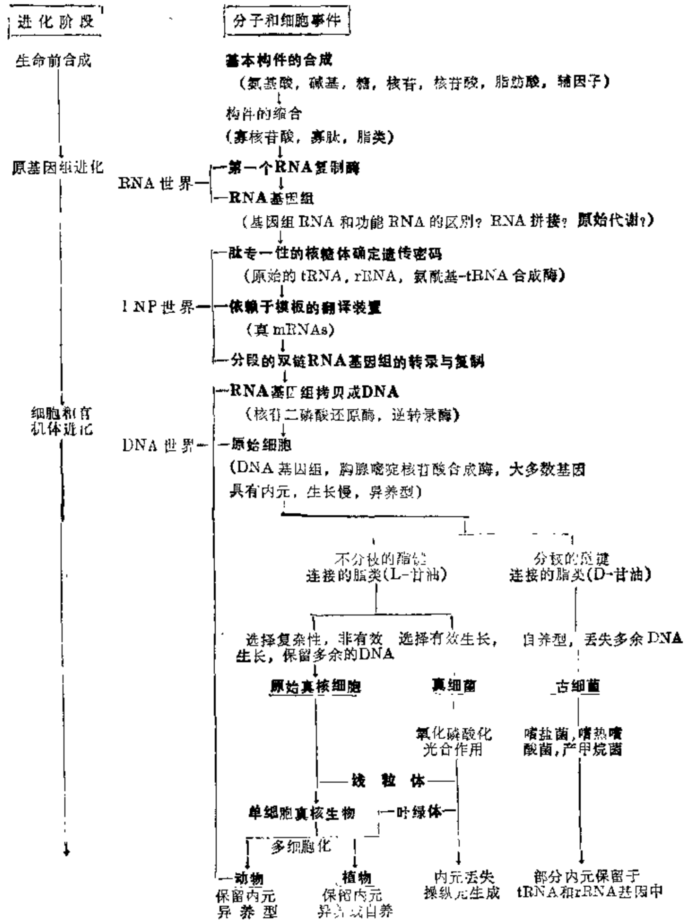



图 2-42　生命的早期进化路线

真核生物的细胞器中最重要的要首推具有遗传物质(DNA)的线粒体和叶绿体了。这两种细胞器是如何产生的？过去有两种假说，一种是内共生假说(endosymbiosis)，一种是分隔假说(segregation)。现在内共生假说得到了广泛的支持(参见第五节)。这种共生关系建立起来以后，内共生体就在真核细胞中繁演下来。真核细胞提供了稳定的环境，内共生体通过氧化磷酸化或光合作用提供能量。长此以往，原线粒体(protomitochondrion)和原叶绿体(protochloroplast)逐渐退化成为细胞器，其一部分DNA消失，一部分转移到核DNA中，只留下少量DNA以编码一些不能透过细胞器膜的分子(rRNA，tRNA，蛋白质)。随着基因数目的减少，细胞器的蛋白质合成机器也发生退化。有人认为，细胞器的遗传密码与通用密码的差异，可能

与蛋白质合成机器的退化有关。但是也有人认为，细胞器的遗传密码代表着较为古老的遗传密码(见第七章)。

线粒体和叶绿体到底是从真细菌的祖先还是从古细菌的祖先演化而来的呢？比较了真核生物线粒体的细胞色素C与细菌的细胞色素C，人们推测线粒体可能来自于紫色光合细菌。然而根据动物线粒体16 S rRNA序列的比较，无法得出结论，因为动物线粒体rRNA序列变化太多，无法得到有意义的比较。幸运的是，植物线粒体16 S rRNA歧化得不多，通过比较得到一个惊人的结果：植物线粒体起源于一组紫色真细菌，包括根瘤菌，土壤杆菌和立克次氏体。这些真细菌直到今天仍然与真核生物共生。

我们根据以16 S rRNA序列分析为基础的三界系统进化树和内共生假说，加上其他的生物化学证据(见以后各章)，就可以得到生命的早期进化的概貌(图2-42)。然而，这仅是初步的推断，其中某些细节将会随着新证据的发现需要加以修正。

## 提　　要

根据细胞核膜的有无，可将有细胞的生命分为原核生物和真核生物两大类，二者具有明显不同的分子遗传学特征。人们对于大多数原核生物和几乎全部真核生物的染色体的DNA序列只有部分的了解。对于真核生物的巨大的DNA含量，人们还无法用已知功能来进行解释，这就是所谓C值矛盾。

原核生物虽然没有核膜，但其DNA与RNA和蛋白质一起形成一个相对集中的区域即类核。原核生物染色体上功能相关的基因大多组成操纵元结构，其蛋白质基因通常为单拷贝，而RNA基因则是多拷贝的。在噬菌体染色体中有时还出现基因重叠和基因内基因。

真核生物具有复杂的染色体结构，其中DNA所占的比例比原核生物的类核显著下降，而蛋白质含量则显著上升。根据真核生物DNA复性动力学的研究，其DNA上存在着单一序列、轻度重复序列、中度重复序列和高度重复序列。许多高度重复序列DNA在进行超离心时会形成浮力密度不同的所谓卫星DNA。卫星DNA通过跃进式复制和突变积累的交替过程而演化，产生了现今卫星DNA的等级结构。交换固定和基因转换与基因扩增共同维持了串联重复序列的均一性。染色体是由染色质构成；染色质分为常染色质和异染色质两种。染色质的基本组成单位是核小体。每个核小体单位包括200bp左右的DNA和一个组蛋白八聚体以及一个分子的组蛋白H1。位于组蛋白八聚体这一核心颗粒上的两圈DNA称为核心DNA，而其余部分称为连接DNA。组蛋白H1是维持正常的核小体结构所不可缺少的。核小体在DNA上的定位与DNA的固有的可弯曲性有关。DNA在核心颗粒表面的左手走向是真核DNA的负超螺旋的主要来源。每个真核染色体都有一个特殊的富含A/T的着丝点序列，它缺乏核小体结构，而与蛋白质结合成为动基体以便与纺锤体的微管蛋白相结合，从而促进复制的染色体在子代细胞中的分配。每个染色体的两端存在着具有正向重复序列的端粒结构，并有特殊的蛋白质与之结合。这可能与真核生物线形DNA末端的复制有关。

大多数真核基因都是不连续基因。在成熟的RNA中出现的部分称为外元，在RNA拼接过程中被删除的部分称为内元。外元和内元的概念与是否编码氨基酸的概念并不相应。通过分子杂交技术与有关技术相结合可以检定不连续基因。真核生物中有许多来源相同、结构相似、功能相关的基因构成基因家族。有些基因家族的成员在染色体上彼此靠近而成串地排列，形成一个基因簇。基因组织形式与基因簇相似的还有串联重复基因，但其序列一致性更高甚

至完全相同，拷贝数也高，且非转录的间隔区短而一致。这类基因的产物都是细胞中大量需要的，如组蛋白基因，rRNA 基因，tRNA 基因等。

在真核细胞中，除了核基因外，在细胞质中也有遗传物质的存在，这就是细胞器基因。动物的线粒体，植物的线粒体和叶绿体都有 DNA，大多为环状，少数为线状。一个细胞中常有许多细胞器 DNA，甚至是同一个细胞器中(如叶绿体)有时也含有一种以上的 DNA 分子。细胞器 DNA 编码自身需要的某些蛋白质以及 tRNA 和 rRNA，其余的蛋白质均由核基因编码，在细胞质中合成后输入细胞器。线粒体和叶绿体膜均为无孔的双层膜，它一般不允许核酸分子的通过，但少数核酸分子可在蛋白质的保护之下由细胞质输入细胞器。细胞器有自己专用的蛋白质合成机构，尽管其许多组分都是由核基因编码的，但是它们与细胞质的蛋白质合成机构的组分并不相同。与此相一致的是，细胞器密码与通用密码也有不少差别。研究得最清楚的细胞器 DNA 乃是酵母和人类线粒体的 DNA。酵母线粒体的某些基因还含有内元。人类线粒体的基因组织形式体现了节俭的原则，转录后处理成为其基因表达的中心环节。人类线粒体的许多特点都表明细胞器是由原核生物通过内共生作用而演化出来的。

为了确定基因在染色体上的位置，人们建立了许多基因定位方法。遗传交换定位法和接合定位法可以作为经典方法的代表，染色体步行法和染色体跳跃法可以说是现代的方法。终极的基因定位方法乃是进行染色体 DNA 的序列分析。

基因扩增和非均等交换促进了多拷贝基因的形成，再通过突变积累、基因重排和自然选择等因素形成多成员的基因家族或形成新的基因。交换固定和基因转换维持了基因家族中某些非等位基因的均一性。内元序列由于很少受到选择压力的限制而突变较快，从而限制了非均等交换的频繁发生。内元的存在促进了外元的重排过程。外元的重排使得某种结构域可以出现在许多功能相关和不相关的蛋白质分子中。内元可能与它所在基因一样古老，而今天的原核生物和低等真核生物由于需要迅速的 DNA 复制而失去了大部分或绝大部分的内元。有证据表明，有些内元也是可以扩散的，这种扩散依赖于该内元所编码的转座酶。不同的基因在进化过程中承受着不同的选择压力；同一种基因的不同部位也承受着不同的选择压力。因此，不能用一种基因的进化速度去推测另一种基因的进化时间表。

基于 16S rRNA 的碱基变化，三界系统理论逐渐取代了两界系统理论，从而使古细菌、真细菌和真核生物成为三个平行的分枝。三者的共同祖先就是原始细胞。真细菌的内共生作用参与了真核生物的进化过程。

## 参 考 文 献


1. Deley, J. and Kerster, K. (1975) "Biochemical Evolution in Bacteria." *Comp. Biochem.* 29B; 1.
2. Bachman, B. J. (1983) "Linkage Map of E. coli K-12, ed.7." *Microbiol. Rev.* 47: 180～230.
3. Donis-Keller, H. et al. (1987) "A Genetic Linkage Map of the Human Genome." *Cell* 51: 319～337.
4. Hobot, J. A. et al. (1985) "Shape and Fine Structure of Nucleoid on Sections of Ultrarapidly Frozen and Cryosubstituted Bacteria." *J. Bacteriol.* 162: 960～971.
5. Sanger, F. et al. (1977) "Nucleotide Sequence of Bacteriophage $\phi\times174$ DNA." *Nature* 265: 687～695.
6. Hendrix, R. W. et al. (1983) *Lambda* Ⅱ. Cold Spring Harbor Laboratory.
7. Rouviere-Yaniv, J. and Yaniv, M. (1979) "*E. coli* Binding Protein HU Forms Nucleosome-like Structures with Circular Double Stranded DNA." *Cell* 17: 265～274.
8. Tanaka, I. et al. (1984) "3-Å Resolution Structure of a Protein with Histone-like Properties in Procaryotes." *Nature* 310: 376～381.
9. Earnshaw, W.C. et al. (1985) "Topoisomerase Ⅱ Is a Structural Component of Mitotic Chromosome Scaffolds."

*J. Cell. Biol.* 100: 1706～1715.
10. McGhee, J. D. and Felsenfeld, G. (1980) "Nucleosome Structure." *Ann. Rev. Biochem.* 49: 1115～1156.
11. Kornberg, R. D. and Klug, A. (1981) "The Nucleosome" *Sci. Amer.* 244: 52～79.
12. Burlingame, R. W. et al. (1985) "Crystallographic Structure of the Octameric Histone Core of the Nucleosome at a Resolution of 3.3 Å." *Science* 228: 546～553.
13. Klug, A. et al. (1982) "Helical Periodicity of DNA on and off the Nucleosome as Probed by Nucleases." *Cold Spring Harbor Symp. Quant. Biol.* 47: 285～292.
14. Richmond, T. J. et al. (1984) "Structure of the Nucleosome Core Particle at 7 Å Resolution." *Nature* 311: 532～537.
15. Felsenfeld G. et al. (1986) "Structure of the 30 nm Chromatin Fiber." *Cell* 44: 375～377.
16. Benezra, R. et al. (1986) "Nucleosomes Are Phased along the Mouse β-major Globin Gene in Erythroid and Non-erythroid Cells." *Cell* 44: 697～704.
17. Stillman, B. (1986) "Chromatin Assembly During SV40 DNA Replication in Vitro." *Cell* 45: 555～565.
18. Travers, A.A. (1987) "DNA Bending and Nucleosome Positioning." *TIBS* 12: 108～112.
19. Carbon, J. (1984) "Yeast Centromeres: Structure and Function." *Cell* 37: 351～353.
20. Murray, A.W. et al. (1986) "Chromosome Length Controls Mitotic Chromosome Segregation in Yeast." *Cell* 45: 529～536.
21. Murray, A. and Szostak, J.W. (1983) "Construction of Artifical Chromosomes in Yeast." *Nature* 305: 189～193.
22. Blackburn, E. (1984) "Telomeres: Do the Ends Justify the Means?" *Cell* 37: 7～8.
23. Shampay, J. et al. (1984) "DNA Sequences of Telomeres Maintained in Yeast." *Nature* 310: 154～157.
24. Gottschling, D.E. and Zakian, V.A. (1986) "Telomere Proteins: Specific Recognition and Protection of the Natural Termini of *Oxytricha* Macronuclear DNA." *Cell* 47: 195～205.
25. Morin, G. B. and Cech, T. R. (1986) "The Telomeres of the Linear Mitochondrial DNA of *Tetrahymena thermophila* Consist of 53 bp Tandem Repeats." *Cell* 46: 873～883.
26. Henderson, E. et al. (1987) "Telomeric DNA Oligonucleotides Form Novel Intramolecular Structures Containing Guanine-guanine Base Pairs." *Cell* 51: 899～908.
27. Greider, C.W. and Blackburn, E. H. (1987) "The Telomere Terminal Transferase of *Tetrahymena* Is a Ribonucleoprotein Enzyme with Two Kinds of Primer Specificity." *Cell* 51: 887～898.
28. Breathnach, R. et al. (1977) "Ovalbumin Gene Is Split in Chicken DNA." *Nature* 270: 314～319.
29. Doel, M. T. et al. (1977) "The Presence of Ovalbumin mRNA Coding Sequences in Multiple Restriction Fragments of Chicken DNA." *Nucleic Acid Res.* 4: 3701～3713.
30. Brack, C. and Tonegawa, S. (1977) "Variable and Constant Parts of the Immunoglobulin Light Chain Gene of a Mouse Myeloma Cell Are 1250 Nontranslated Bases Apart." *Proc. Nat. Acad. Sci.* 75: 2205～2209.
31. Jeffreys, A. J. and Flavell, R.A. (1977) "The Rabbit Beta-globin Gene Contains a Large Large Insert In the Coding Sequences." *Cell* 12: 1097～1108.
32. Breathnach, R. and Chambon, P. (1981) "Organization and Expression of Eucaryotic Split Genes Coding for Proteins." *Ann. Rev. Biochem.* 50: 349～383.
33. Chambon, P. (1981) "Split Genes." *Sci. Amer.* 244: 60～71.
34. Church, G. et al. (1980) "Yeast Mitochondrial Intron Products Required in Trans for RNA Splicing." in *Mobilization and Reassembly of Genetic Information* (edited by D.R. Joseph et al. Acad. Press). pp.379～395.
35. Michel, F. et al. (1985) "Mitochondrial Class II Introns Encode Proteins Related to the Reverse Transcriptases of Retroviruses." *Nature* 316: 641～643.
36. Carignani, G. et al. (1983) "An mRNA Maturase Is Encoded by the First Intron of the Mitochondrial Gene for the Subunit I of the Cytochrome Oxidase in *S. Cerevisiae*." *Cell* 35: 733～742.
37. Colleaux, L. et al. (1986) "Universal Code Equivalent of a Yeast Mitochondrial Intron Reading Frame Is Expressed into *E. coli* as a Specific Double Strand Endonuclease." *Cell* 44: 521～533.
38. Riccio, A. et al. (1985) "The Human Urokinase—plasminogen Activator Gene and Its Promoter." *Nucl. Acid Res.* 13: 2759～2771.
39. Van der Ploeg, L. H. T. (1986) "Discontinuous Transcription and Splicing in *Trypanosomes*." *Cell* 47: 479～480.
40. Chu, F.K. et al. (1986) "Characterization of the Intron in the Phage T4 Thymidylate Synthase Gene and Evidence for Its Self-excision from the Primary Transcript." *Cell* 45: 157～166.
41. Yamada, Y. et al. (1980) "The Collagen Gene, Evidence for Its Evolutionary Assembly by Amplification of a

DNA Segment Containing an Exon of 54 bp." *Cell* 22: 887～892.
42. Maniatis, T. et al. (1982) "Molecular Genetics of Human Globin Gene Expression." in *Molecular Genetics of Neuroscience*, ed. Schmitt, F.O. et al., New York, Raven Press, pp. 87～101.
43. Maxson, R. et al. (1983) "Expression and Organization of Histone Genes." *Ann. Rev. Genetics* 17: 239～277.
44. Hentschel, C.C. and Birnstiel, M.L. (1981) "The Organization and Expression of Histone Gene Families." *Cell* 25: 301～313.
45. Long, E. O. and Dawid, I.B. (1980) "Repeated Genes in Eucaryotes." *Ann. Rev. Biochem.* 49: 727～764.
46. Planta, R. J. and Meyerink, J. H. (1980) "Organization of the Ribosomal RNA Genes in Eucaryotes." in *Ribosomes: Structure, Function and Genes*, ed. Chambliss, G. et al., Baltimore: University Park Press, pp. 871～887.
47. Clarkson, S.G. (1983) "Transfer RNA Genes." in *Eucaryotic Genes: Their Structure, Activity and Regulation*, ed. Maclean, N. et al., pp.239～261.
48. Anderson, S. et al. (1981) "Sequence and Organization of the Human Mitochondrial Genome." *Nature* 290: 457～465.
49. Attardi, G. (1985) "Animal Mitochondrial DNA: An Extreme Example of Genetic Economy." *Internatl. Rev cytol.* 93: 93～145.
50. Chomyn, A. et al. (1986) "URF6, Last Unidentified Reading Frame of Human mtDNA, Codes for An NADH Dehydrogenase Subunit." *Science* 234: 614～618.
51. Chang, D.D. and Clayton, D. A. (1987) "A Mammalian Mitochondrial RNA Processing Activity Contains Nucleus-Encoded RNA." *Science* 235: 1178～1184.
52. Wollman, E. L. et al. (1956) "Conjugation and Genetic Recombination in *E. coli*." *Cold Spring Harbor Symp. Quant. Biol.* 21: 141～162.
53. Warren, G. J. et al. (1978) "Col E1 Plasmid Mobility and Relaxation Complex." *Nature* 274: 259～261.
54. Bender, W. et al. (1979) "Gene Isolation by Chromosomal Walking." *J. Supra. Mol. Struc.* 10 (suppl.): 32.
55. Chinault, A.C. and Carbon, J. (1979) "Overlap Hybridization Screening: Isolation and Characterization of Overlapping DNA Fragments Surrounding the LEU 2 Gene on Yeast Chromosome Ⅲ." *Genes* 5: 111～126.
56. Clarke, L. and Carbon, J. (1980) "Isolation of a Yeast Centromere and Construction of Small Circular Chromosomes." *Nature* 287: 504～509.
57. Poustka, A. et al. (1987) "Construction and Use of Human Chromosome Jumping Libraries from Not I Digested DNA." *Nature* 325: 353～355.
58. Go, M. (1981) "Correlation of DNA Exonic Regions with Protein Structural Units in Haemoglobin." *Nature* 291: 90～92.
59. Branden, C.-I. et al. (1984) "Correlation of Exons with Structural Domains in Alcohol Dehydrogenase." *EMBO J.* 3: 1307～1310.
60. Gilbert, W. et al. (1986) "On the Antiquity of Introns." *Cell* 46: 151～154.
61. Südhof, T.C. et al. (1985) "Cassette of Eight Exons Shared by Genes for LDL Receptor and EGF Precursor." *Science* 228: 893～895.
62. Marchionni, M. and Gilbert, W. (1986) "The Triosephosphate Isomerase Gene From Maize: Intron Antedate the Plant-animal Divergence." *Cell* 46: 133～141.
63. McKnight, G.L. et al. (1986) "Nucleotide Sequence of the Triosephosphate Isomerase Gene from *Aspergillus nidulans*: Implications for a Differential Loss of Introns." *Cell* 46: 143～147.
64. Littman, D.R. et al. (1987) "Unusual Intron in the Immunoglobulin Domain of the Newly Isolated Murine CD4 (L3T4) Gene." *Nature* 325: 453～455.
65. Weiner, A. M. and Denison, R.A. (1982) "Either Gene Amplification or Gene Conversion May Maintain the Homogeneity of the Multigene Family Encoding Human U1 Small Nuclear RNA." *Cold Spring Harbor Symp. Quant. Biol.* 47: 1141～1149.
66. Jaworski, C.J. and Piatigorsky, J. (1989) "A Pseudo-Exon in the Functional Human αA-Crystallin Gene." *Nature* 337: 752～754.
67. Woese, C.R. (1981) "Archaebacteria." *Sci. Amer.* 244: 98～125.
68. Pace, N. R. et al. (1986) "Ribosomal RNA Phylogeny and the Primary Lines of Evolutionary Descent." *Cell* 45: 325～326.
69. Kjems, J. and Garrett, R.A. (1985) "An Intron in the 23S Ribosomal RNA Gene of the Archaebacterium *Desulfurococcus mobilis*." *Nature* 318: 675～677.
70. Yang, D. et al. (1985) "Mitochondrial Origins." *Proc. Nat. Acad. Sci.* 82: 4443～4447.

71. Gott, J.M. et al. (1986) "Multiple Self-Splicing Introns in Bacteriophage T4, Evidence from Autocatalytic GTP Labeling of RNA in Vitro." *Cell* **47**: 81～87.
72. Sjöberg, B.-M. et al. (1986) "The Bacteriophage T4 Gene for the Small Subunit of Ribonucleotide Reductase Contains an Intron." *EMBO J*. **5**: 2031～2036.
73. Michel, F. and Dujon, B. (1986) "Genetic Exchanges between Bacteriophage T4 and Filamentous Fungi?" *Cell* **46**: 323.
74. Cavalier-Smith, T. (1989) "Archaebacteria and Archezoa." *Nature* **339**: 100～101.
75. Gouy, M. and Li, W.-H. (1989) "Phylogenetic Analysis Based on rRNA Sequences Supports the Archaebacterial Rather than the Eocyte Tree." *Nature* **339**: 145～147.
76. Joyce, G.F. (1989) "RNA Evolution and the Origins of Life." *Nature* **338**: 217～224.
77. Pool, R. (1990) "The Third Kingdom of Life." *Science* **247**: 159.
78. Pool, R. (1990) "Pushing the Envelope of Life." *Science* **247**: 158～160.

# 第三章 DNA 的复制

从 Watson 和 Crick 建立 DNA 结构的双螺旋模型开始，关于 DNA 复制原理的轮廓就已经诞生，正如他们二人给《Nature》杂志的第一封信中所写："我们并没有忽视，从我们所假设的特异的配对方式可以提出一个关于遗传物质可能的复制机制的建议来"。然而，实际的过程却远为复杂，至今人们对于复杂的 DNA 复制过程还没有了解得十分清楚。这一过程的复杂性在于：

(1) 复制过程需要能量供应以解开双螺旋链；

(2) 已经解开的单链也可能产生链内碱基配对；

(3) 一种酶只能催化有限的物理化学反应；

(4) 复制过程必须设计若干安全保障以防止复制的错误，并纠正业已发生的错误(尽管极少)；

(5) 巨大的DNA分子特别是环形分子给复制带来许多几何学的问题，比如 *E.coli* 复制速度为 $10^5$bp/分，如果 DNA 模板以此速度解旋的话，则相当于每小时开 70 英里的汽车引擎转动的速度；

(6) 对于所有含有双螺旋DNA的生物有机体，并没有单一的可以普遍适用的复制机制。然而，各种生物的 DNA 复制还是有共同点的，例如新生的 DNA 链总是按 A·T，G·C 这样的碱基配对规律与母体链互补，新生的 DNA 链的单体总是由 DNA 聚合酶一个一个地加在这条新生链的 3′-OH 末端。

## 第一节 DNA的半保留复制[1,2]

Watson 和 Crick 最早提出 DNA 的半保留复制机理，就是在复制过程中各以双螺旋 DNA 的其中一条链为模板合成其互补链，新生的互补链与母链构成子代 DNA 分子。

这一假说于 1957 年得到 Matthew Meselson和 Franklin Stahl 所设计的精巧的实验所证实。为了节省篇幅，我们仅用图解加以说明(图 3—1)。

这一实验有力地支持了半保留复制的假说，它不但否定了全保留复制假说，而且也排除了弥散复制假说。所谓弥散复制就是说子代 DNA 的每一条链都是由亲本链的片断与新合成的片断随机拼接而成。

## 第二节 复制原点、方向和方式[3~6]

很多实验都证明了复制是从 DNA 分子上的特定位置开始的，这一位置叫复制原点，常用 *ori* 或 *o* 表示。例如大肠杆菌的复制原点位于 *ilv* 基因附近，是一个包括大约 245 bp 的一个区段。现已证明，除 fd 组的噬菌体以外，许多生物的复制原点都是双螺旋 DNA 呼吸作用强烈的区段，即经常开放的区段(frequently opening region)，亦即富含 A·T 的区段。这一区

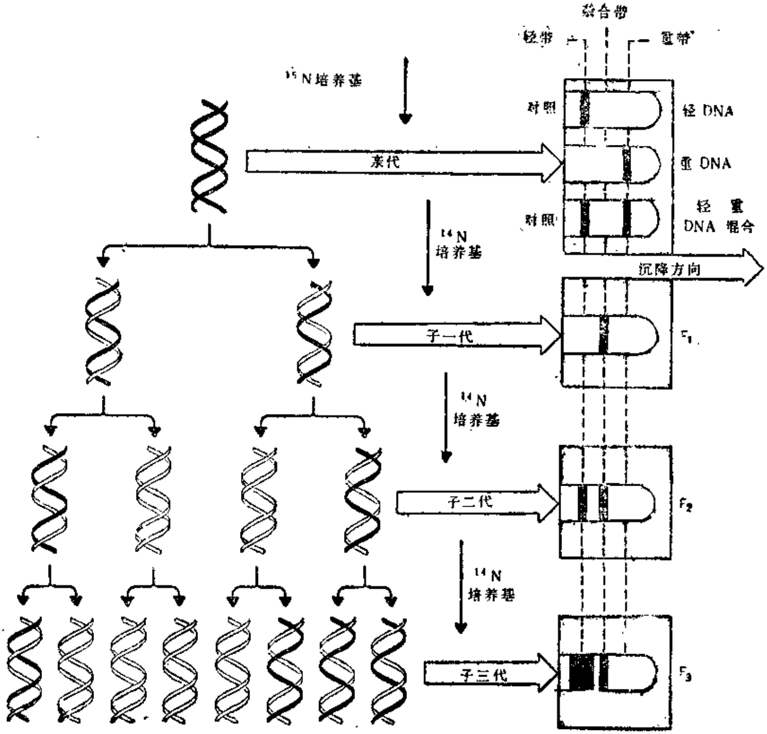



图 3-1 Meselson-Stahl实验

支持半保留复制假说的 CsCl 超离心试验结果显示不同比重的 DNA 带

段产生的瞬时单链与 ssb 蛋白 (single-stranded DNA binding protein) 结合，对复制的起始十分重要。在所有的原核生物中，复制都是从特定的位置开始的，迄今尚未发现例外。

DNA 复制从特定位置开始，大多数双向进行，也有一些单向的，或以不对称的双向方式进行的。这是由复制原点的性质所决定的。

用遗传学方法可以证明 *E.coli* 的复制是从 *ilv* 基因附近开始，以双向等速进行复制的。细菌染色体复制一次的时间相当于细胞繁殖一代所需要的时间。利用温度突变种能使 DNA 复制同步进行。这种突变在 25℃ 时能正常繁殖，而在 42℃ 时虽能完成复制之中的 DNA 合成，但不能开始新的一轮复制。如果每个细菌从原点同时开始复制，在一定时刻，靠近复制原点的基因复制得多些，而远离复制原点的基因就复制得少些。假如复制是单向进行的，基因频率在基因图上呈单向梯度，最高频率的基因和最低频率的基因应该是连锁的(因为是双链环状染色体)。如果是双向复制，基因频率将以 O 点为中心出现双向梯度。测定了 *E.coli* 的许多

基因频率，发现是以 *ilv* 为中心双向下降。这说明 *E.coli* 的复制原点在 *ilv* 基因附近。关于 *E.coli* 复制原点的性质将在第六节中叙述。

噬菌体 $\mu$ 能插入大肠杆菌染色体上不同的位置，而噬菌体 $\lambda$ 只能插入一个特定的位置。在对数生长期可以测定噬菌体 $\mu$ 和 $\lambda$ 的数量。实验结果表明，噬菌体 $\mu$ 插入到 *ilv* 附近，不论是在右边或左边，相对产量都是最高，而插入 *ilv* 的 180° 的位置，产量最低。从这个实验可以得出结论：*E.coli* DNA 的复制从 *ilv* 基因开始，双向等速进行。用放射自显影方法也可以得到有关复制原点和方向的资料。

用电子显微镜观察噬菌体T7DNA复制，总是在离一端 17% 处出现一个复制眼，然后向两边伸展。实验表明，真核生物 DNA 的复制也是从特定位置开始，以双向等速进行，只不过是一个 DNA 分子上有许多个特定的复制原点。

也有一些例外的情况。例如在枯草杆菌中，复制从原点开始双向复制，但两个复制叉的移动不对称，一个移动 1/5 的距离便停下来，然后另一个复制叉走完 4/5 的距离。

质粒 R6K 复制的早期是单向进行的，可是这一复制叉在离起点 1/5 处停下来，然后从相反的方向启动第二个复制叉。

线粒体 DNA 的复制也是不对称的，DNA 双链中的一条先复制成双链环态，而另一条亲本链被置换出来成为单链状态，叫做置换噌噗，又叫 D噌噗(displacement loop)。一条链复制 67%，另一条链才开始复制。这一复制方式的生物学意义目前还不太清楚。

质粒 *ColE*1 的复制完全是单向的。从体内或体外系统中分离到的复制中间体，其一个"复制叉"(其实就是复制原点)总是离开 EcoRI 切点 17% 左右，而另一个复制叉离开 EcoRI) 切点的距离却是可变的。因此 *ColE* 1 的复制是从特定位置开始，单方向进行。

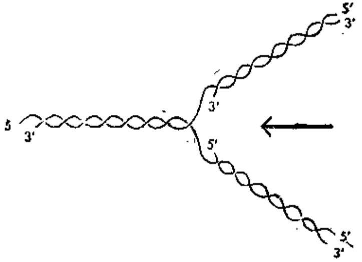



图3-2 复制叉 箭头表示子链总的生长方向

根据 DNA 合成的起始方式，复制可以分为两种类型，一类叫从新起始(de novo initiation)，即先导链是从新开始的，这种类型主要是复制叉式复制；另一类叫共价延伸，即先导链是共价结合在一条亲本链上，这一类型主要是滚环式复制。

(1) 复制叉式(包括 $\theta$ 复制)

DNA 在复制原点解开成单链状态，分别作为模板，各自合成其互补链，则出现两个叉子状的生长点，叫做复制叉(图 3-2)。具有双链环状 DNA 的细菌和病毒如何复制其DNA呢？John Cairns 做了卓越的实验。他用 $^3H$ 标记的胸腺嘧啶培养 *E.Coli*，通过放射自显影得到结果如图 3-3 所示。由于形状像希腊字母 $\theta$，因而叫 $\theta$ 复制，又叫 Cairns 复制，其复制中间体称为 Cairns 分子。在 $\theta$ 复制中同样有双向复制和单向复制(图 3-4)，大多数生物为双向等速复制，也有少数双向不等速复制的情况。

(2) 滚环式复制

双链环状 DNA 除了 $\theta$ 复制之外，某些双链环状 DNA(或以某种方式变为双链环状 DNA) 在复制时，以某种方式切断其中一条链。其 5′ 端常与特殊的蛋白质相连，而在 3′ 端不断地由 DNA 聚合酶催化，以未切断的一条环形链为模板，加上新的核苷酸。由于 3′ 端不断延长，而 5′ 端则不断地甩出。好像中间的一个环在不断地滚动一样，因而叫做滚环复制。因其形状像

图 3-3 *E.coli* 的 $\theta$-复制中间体的电子显微镜照片及解释图，A，B 两部分已复制，X，Y 表示两个复制叉。

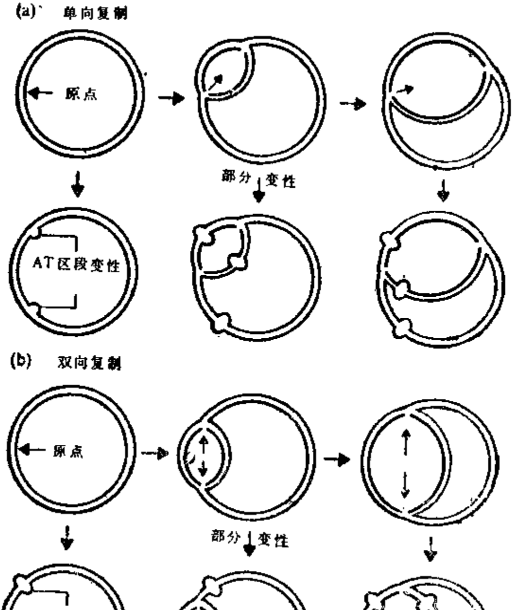



图 3-4 单向 $\theta$ 复制和双向 $\theta$ 复制模式图。该实验以两个富含 A-T 区段为指示，在 pH11.05 和 25℃ 下让 DNA 部分变性 10 分钟，从而显示出复制叉与变性的 A-T 区段的相对距离。

希腊字母 $\sigma$，因此又叫 $\sigma$ 复制(图 3-5)。当这条链甩到某一定长度则开始复制其互补链。这样，常常会得到一种具有原来 DNA 分子的若干倍的中间产物，叫做连环分子 (Concatemer)。如图 3-6 示。

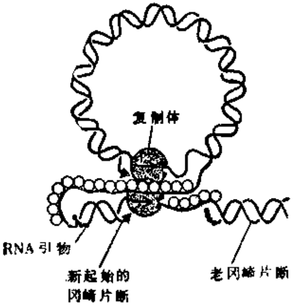



图 3-5 滚环复制的一般模式

滚环复制在某些噬菌体 DNA 的复制、细菌交配和真核生物的基因扩增中占有重要位元。在讲过复制的酶学机制后，我们再进一步阐述滚环复制。

…XYZABC……XYZABC……XYZABC……XYZABC…

图 3-6 滚环复制的中间产物连环分子(concatemer)

## 第三节 DNA 复制的酶学[7~19]

DNA 复制是一个非常复杂的酶学过程，它需要 30 种以上的酶和蛋白质，人们常把这些酶和蛋白质称为复制体系，其中一部分酶和蛋白质结合在一起，协同动作，构成所谓复制体 (replisome)。

### 一、DNA 的聚合反应和聚合酶

1957 年，Arthur Kornberg 首次发现 DNA 聚合酶 I，后来又相继发现了 DNA 聚合酶 Ⅱ、DNA 聚合酶 Ⅲ。DNA 聚合酶 Ⅱ 的作用还不清楚，根据实验室工作，它在 DNA 复制过程中似乎没有什么作用，DNA 聚合酶 Ⅲ 在 *E.coli* 中才是主要的聚合酶。以后，三种 DNA 聚合酶分别简写为 Pol I、Pol Ⅱ、Pol Ⅲ。

DNA 聚合酶只能以四种核苷 5′-三磷酸为前体，即 dATP、dGTP、dCTP、dTTP，而不能是 5′-二磷酸，5′-单磷酸，也不能是 3′-三磷酸，3′-二磷酸、3′-单磷酸。反应通式可以写置：

$$\text{Poly (核苷酸)}_n\text{-3}'\text{-OH} + \text{dNTP} \longrightarrow \text{Poly(核苷酸)}_{n+1}\text{-3}'\text{-OH} + 2\,\text{Pi}$$

这里 Pi 是从焦磷酸分解而成的无机磷酸。

由于多核苷酸的磷酸二酯键水解时释放出大量的能量，自由能为负值，因而单磷酸核苷就不可能加到生长的 DNA 链的末端。实际上，磷酸二酯键能自发地水解，尽管其速度很低。dNTP 和 DNA 链发生反应时，虽然有一个高能键水解，但对于聚合反应来说，总的自由能仅

仅略大于 0(+0.5 千卡/克分子≈2100 焦耳/克分子)，因而很难使反应向聚合方向转移。为了驱动这一反应向聚合方向转移，焦磷酸就必须水解为无机磷酸。这一反应由焦磷酸酶催化，反应释放出大量的负自由能，从而保证了 DNA 聚合反应的迅速进行。这也解释了为什么不能以核苷二磷酸作为前体来合成 DNA。

## 二、脱氧核苷三磷酸前体的来源

脱氧核苷三磷酸前体有两个来源，一个是废物利用途径，即体内核酸降解的产物或直接来自培养基中；另一个是从新合成的途径下面只叙述从新合成途径。

生物体可以利用氨基酸，$CO_2$，$NH_3$和PRPP(磷酸核糖焦磷酸，phosphoribosylpyrophosphate) 合成核糖核苷单磷酸 (具体反应请参考生物化学教程)。然后，核糖核苷单磷酸(NMP)在各自的具有碱基特异性的核苷单磷酸激酶 (没有核糖和脱氧核糖的特异性) 的作用下转变成NDP：

$$NMP + ATP \longrightarrow NDP + ADP$$

接着，在一种没有碱基特异性的核糖核苷酸还原酶的催化下，脱掉核糖 2′ 位上的氧：

$$NDP \longrightarrow dNDP$$

然后，由一种既无碱基特异性又无糖环特异性的核苷二磷酸激酶将 dNDP 转变成 dNTP：

$$dNDP + ATP \longrightarrow dNTP + ADP$$

这里的 dNTP 包括 dATP、dGTP、dCTP、dUTP。

必需注意，合成 DNA 并不需要 dUTP，合成 RNA 也不需要 dUTP。生物体内有一系列防范 dUTP 进入 DNA 的措施，一方面 dUTP 在 dUTP 酶的作用下可以分解为 dUMP 和 PPi；另一方面，万一 dUTP 渗入 DNA 分子中，可以通过尿嘧啶 DNA 糖基酶而得到修复(详见修复一章)。

合成 DNA 所不可缺少的 TTP(即 dTTP))是这样生成的：

$$dUTP \xrightarrow[\searrow PPi]{dUTPase} dUMP + \text{四氢叶酸} \xrightarrow{\text{胸腺嘧啶核苷酸合成酶}} TMP + \text{二氢叶酸}$$

式中的四氢叶酸是甲基供体。然后 TMP 在胸腺嘧啶核苷酸激酶作用下生成 TDP，接着，核苷二磷酸激酶把 TDP 变为 TTP。

胸腺嘧啶核苷酸合成酶 (thymidylate synthetase) 在研究 DNA 复制机制时，需要使之缺乏或受抑制，以便让生物体只能利用培养基中的标记的胸腺嘧啶或胸腺嘧啶核苷。在微生物中可以找到缺乏此酶的突变体 (*thy*⁻)；在真核生物中可以加氟尿嘧啶脱氧核苷 (FUDR，flurouracil deoxyriboside) 以抑制该酶。FUDR 在体内磷酸化以后与该酶和四氢叶酸生成一个三元复合物，从而抑制了酶活性。

## 三、三种 DNA 聚合酶的结构和功能

我们用列表的方式说明 *E.coli* 的三种 DNA 聚合酶的基本特征(表 3-1)。DNA 聚合酶Ⅲ极为复杂，目前已知它的全酶包含 10 种共 22 个亚基，其组成是 $\alpha_2\varepsilon_2\theta_2\tau_2\gamma_2\delta_2\delta'_2\chi_2\psi_2\beta_4$。DNA 聚合酶Ⅲ的成分列于表 3-2，它们可能的装配模型示于图 3-7。

表 3-1 大肠杆菌 DNA 聚合酶特性

| | DNA 聚合酶Ⅰ | DNA 聚合酶 Ⅱ | DNA 聚合酶 Ⅲ |
|---|---|---|---|
| 分子结构 | 单链分子 | 单链分子 | 核心酶有 3 个亚基<br>全酶有 22 个亚基 |
| 分子量(KDa) | 109 | 90 | 900 |
| 分子数/细胞 | 400 | 100 | 10 |
| 5'→3' 聚合酶活性 | + | + | + |
| 37℃时的转化率<br>(核苷酸数/酶分子·分) | ～600 | ～30 | ～9 000 |
| 3'→5' 外切核酸酶活性 | + | + | + |
| 5'→3' 外切核酸酶活性 | + | − | + |
| 内切核酸酶活性 | + | − | − |
| 切刻转移活性 | + | − | − |
| 对核苷三磷酸的亲和力 | 低 | 低 | 高 |
| 高盐浓度抑制 | − | − | + |
| SH 封闭剂抑制 | − | + | + |

表 3-2 DNA 聚合酶 Ⅲ 全酶的亚基及亚聚集体

| 亚基 | 基因 | 质量(KDa) | 功能及组成的亚聚集体 | | | |
|---|---|---|---|---|---|---|
| α | *dnaE* | 132 | 聚合酶 | 校对功能 | Pol Ⅲ 核心酶 165 KDa | Pol Ⅲ 410 KDa | Pol Ⅲ* 800 KDa |
| ε | *dnaQ* | 27 | 3'→5' 外切酶 | | | | |
| θ | | 10 | 核心酶相互连接(待证明) | | | | |
| τ | *dnaZX* | 71 | 将核心酶与模板结合，ATPase 活性 | | | | |
| γ | *dnaZ(X)* | 52 | 组成 γ 复合体 | | | | |
| δ | | 35 | 组成 γ 复合体；又称因子Ⅲ；过去误认为 *dnax* 产物 | | | | |
| δ' | | 33 | 组成 γ 复合体 | | | | |
| χ | | 15 | 组成 γ 复合体 | | | | |
| ψ | | 12 | 组成 γ 复合体 | | | | |
| β | *dnaN* | 37 | 使全酶具有进行性 | | | | |

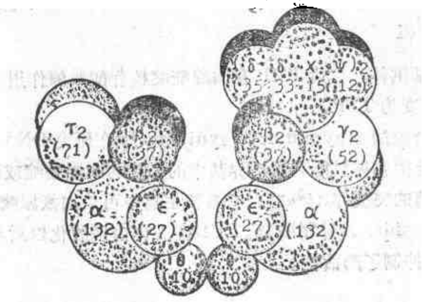



图 3-7 DNA 聚合酶 Ⅲ 全酶的暂定的装配模型

现将 DNA 聚合酶Ⅰ和Ⅲ的主要特性和功能介绍如下：

**1. DNA 聚合酶活性** polⅠ，polⅢ 都具有 DNA 聚合酶活性，而且都需要模板和引物，只有当引物具有自由的 3'-OH 时才能将四种前体核苷酸按碱基配对规律合成 DNA 链。但

polI 的聚合酶活性主要用 DNA 的修复和 RNA 引物的替换；polⅢ 才是使DNA 链延长的主要的聚合酶。

2．3′→5′外切核酸酶活性　polI，polⅢ 都有这种酶活性，这种酶活性是基于对不配对碱基造成的单链的识别。因此这种酶活性是保证其聚合作用的正确性不可缺少的，这种功能称为校对功能（也有称为编辑功能的），这对于 DNA 作为遗传物质所必需的稳定性和极高的保真度是至关重要的。其作用原理见图 3-8 所示。

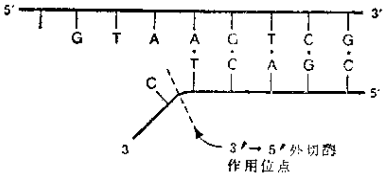



图 3-8　DNA聚合酶的3′→5′外切核酸酶活性。箭头所指为外切酶作用位点。

3．5′→3′外切核酸酶活性　polI 和 polⅢ 的这种酶活性，相互之间有很大的区别。PolⅢ 的5′→3′外切核酸酶活性只能作用于单链 DNA，因此，它不能产生切刻平移作用 (nick translation)。polⅢ 的 5′→3′外切核酸酶活性的生物学意义还不知道。

PolI 的 5′→3′ 外切核酸酶活性有如下特征：

(1) 核酸必须有 5′-磷酸末端；在切刻或缺口处，只要有 5′-P 末端，这种酶活性同样可以发挥出来。

(2) 被去除的核苷酸必须是已经配对的；以一个接着一个的方式去除。

(3) 被去除的核苷酸可以是脱氧核糖核苷酸，也可以是核糖核苷酸。图 3-9 表示 PolI 的 5′-3′ 外切酶活性的若干类型底物和产物。

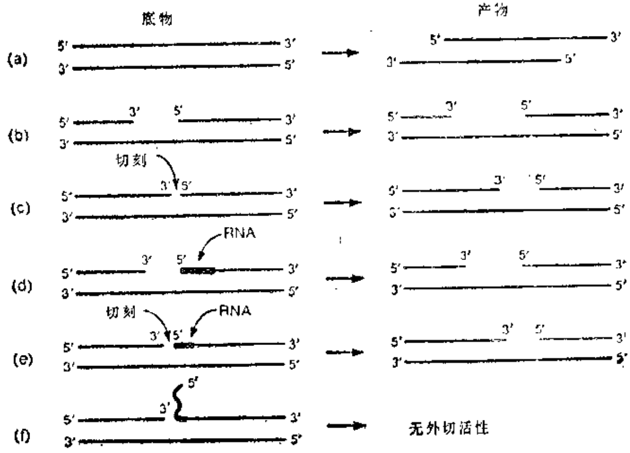



图 3-9　polI 的 5′→3′ 外切酶活性示意图

(a) 两个 5′末端被去除　(b) 缺口的扩大　(c) 切刻转变成缺口　(d) 缺口处具有 5′-P 的 RNA 被去除　(e) 切刻处具有 5′-P 的 RNA 被去除　(f) 对没有配对的 5′-P 末端，外切酶活性不起作用　而代之以内切酶活性(见图 3-12)

下面讲几个与 PolI 的 5′→′3 外切酶活性和聚合酶活性有关的现象，一个是切刻平移(nick translation)，一个是链的置换，另一个是模板转辙(template switching)。

在切刻的 5′→P 末端由 5′→3′ 外切酶活性进行水解反应，而同时在切刻的 3′-OH 末端由聚合酶活性进行聚合反应。由于 polI 不能像 DNA 连接酶那样共价连接相邻的 3′-OH 末端和 5′-P 末端，因而上面原有的切刻仍然是切刻，只不过是随着两种反应的进行，由 5′ 端向 3′ 端转移，这种现象叫做切刻平移(见图 3-10)。

在实验条件下，可以让切刻的 3′-OH 末端进行聚合反应而不让切刻的 5′-P 末端发生外切酶的水解反应。这样，生长的链就置换了亲本链。目前已知的 *E.coli* 三种 DNA 聚合酶中，只有 polI 才能进行这种不需辅助的置换反应。在其他一些链的置换反应中，需要辅助性的蛋白质并需要水解 ATP 以供应解开螺旋所需的能量。链置换反应亦示于图 3-10。

在体外实验中，常常与链的置换作用相伴随的是一种奇怪的反应，即随着置换作用的进展，生长链抛弃了原来的模板，而转靠被置换出来的亲本链为模板，这种现象叫模板转辙(图3-11)。模板转辙可能只是实验条件下产生的现象，或许没有什么生物学意义。

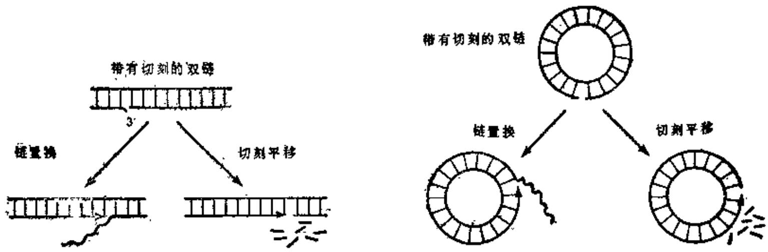



图 3-10　切刻平移和链的置换

图 3-11　模板转辙示意图

4. 核酸内切酶活性　当一段 DNA 带有 5′-P 的不配对单链时，polI 可以作用在与此单链相连的两个配对碱基之间，切断其磷酸二酯键，如图 3-12 所示。

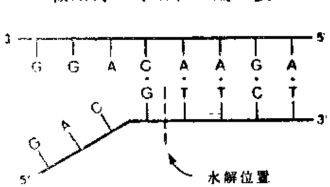



图 3-12　polI 的内切核酸酶的活性

5. 缺口充填能力　polⅢ 和 polI 缺口充填能力是大不相同的。polⅢ 只能充填几个碱基的小缺口，而 polI 能充填很大的缺口，甚至能完成几乎整个互补链的复制。这实际上是反映了 polI 和 polⅢ 的聚合酶活性所要求的条件的不同。图 3-13 表示了几种缺口充填的类型。

6. DNA 聚合反应进行的方向　已知的 DNA 聚合酶只能使链按 5′→3′ 方向生长，RNA

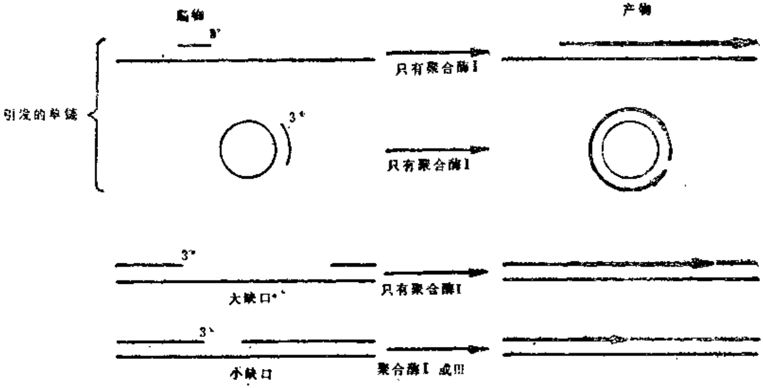



图 3-13　polI 和 polIII 的缺口充填能力

聚合酶合成 RNA 时也是如此。生物在进化中选择这样的方式是有它的深刻的化学上和功能上的根源的。假若链的生长按 3′→5′ 的方向进行，则 5′ 端必须带有三磷酸基团才能与下一个前体的 3′-OH 起反应。这个三磷酸基团带有强烈的负电荷，尽管处在生理盐浓度下，但由于它位于新生链的生长点上，对于末端的碱基配对是极为不利的(其原因在第一章中已经讲过)，因而链的生长也是很困难的；要不然，就得有一个复杂的机构来消除这种影响。另一方面，假定能设法消除上面的影响，但当 DNA 聚合酶行使校对功能时，需要切除末端错误的核苷酸，切除后在 5′ 端只剩下一个磷酸基，又必须有额外的能量供应以及别的酶参加反应才能与下一个正确的前体的 3′-OH 生成磷酸二酯键。这样做既浪费能量，又浪费时间，而且使复制体系增加了不必要机构与环节。因此，所有生物的核酸链的合成都是按 5′→3′ 方向进行的，无一例外。

## 四、DNA 连接酶

由图 3-13 可以看到，DNA 聚合酶虽能充填缺口，甚至聚合了很长的 DNA互补链，但都无法使切刻接合。只有 DNA 连接酶可以完成这一项任务。只有当切刻的 3′-OH 和 5′-P 相邻并且各自的碱基处于配对状态时，DNA 连接酶才能起作用(图 3-14)。

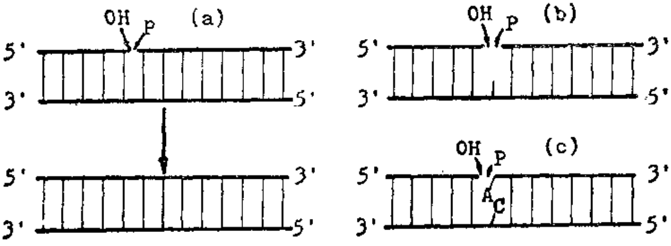



图 3-14　DNA 连接酶要求的底物
底物 (a) 可被连接好，底物 (b) 和 (c) 均不能作用。

由于连接反应是在 3′-OH 和 5′-P 之间进行。因此必须由 ATP 或NAD的水解来供应能量，在真核生物中使用 ATP，而 *E.coli* 连接酶使用 NAD(nicotine adenine dinucletide)。

## 五、与 DNA 几何学性质相关的酶

1. 螺旋酶 (helicases) 又称为解旋酶，是促进 DNA 的两条互补链分离的一类酶，细胞中含有多种螺旋酶。它们的功能相互重迭或平行。这大概是由于许多反应都需要 DNA 螺旋的分离，特别是 DNA 的复制这个精确度要求特高的反应更是如此。绝大多数螺旋酶在解链过程中都需要水解 ATP 供应能量，它们具有依赖于 DNA 的 ATPase 活性。

螺旋酶可以分为两类：一类是反应后酶分子按化学计量方式与产物单链 DNA 结合，如螺旋酶 I，螺旋酶 II；另一类是仅仅催化 DNA 双链的分离，它们常与单链 DNA 结合蛋白(简称 SSB 蛋白，即 single-stranded DNA binding protein)配合，以防止链间或链内"退火"，例如 *E.coli* 的 rep 蛋白，螺旋酶Ⅲ，$T_7$ 的基因 4 蛋白，$T_4$ 的 46/61 与 45 的复合物等。人们相信，在 *E.coli* 中 rep 蛋白是最主要的螺旋酶，它解开一个碱基对，需要水解 2 个 ATP 分子。

2. DNA 旋转酶 (Eco 拓扑异构酶 II)：在第一章中我们已经提到过这一重要的拓扑异构酶。

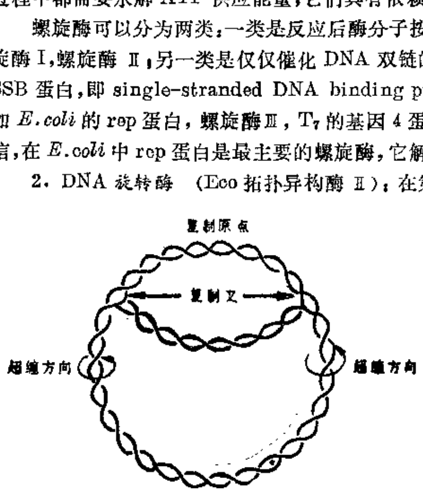



图 3-15 复制叉的移动带来的拓扑学问题

在第二章中我们还讲到，大多数天然 DNA 都具有负超螺旋，它不但可以变为泡状单链形式，有利于许多蛋白质的结合，而且在复制的开始阶段，可以缓解由于复制叉的移动而造成的超缠现象。然而，当 5% 的环形 DNA 双链已经复制时，原有的负超螺旋就被用尽，如果复制叉再向前进，就产生了拓扑学问题(图 3-15)。

这个问题是由拓扑异构酶来解决的。在*E.coli* 中，主要就是DNA旋转酶(Eco 拓扑异构酶 II)，它能产生负超螺旋并消除复制叉移动所产生的正超螺旋。如果加入几种DNA 旋转酶的抑制剂，如香豆霉素(Coumermycin)，新生霉素(novobiocin)，萘啶酮酸(nalidixic acid)等均能抑制细菌 DNA 的合成。在噬菌体感染的细胞中，这些试剂能防止噬菌体 DNA 的超螺旋化，例如 λ 噬菌体 DNA 只能环化而不能超螺旋化，最终形成 $I^*$ 形 DNA。另外，从野生型菌株中分离出来的螺旋酶能结合这些抑制剂，而从能抵抗这些抑制剂作用的突变体中分离出来的旋转酶则不能结合这些抑制剂。

探求 DNA 旋转酶作用机制的基本线索是：该酶对 DNA 分子的每一次作用总是引入两个负超螺旋，因此在一个特定 DNA 分子中，它的超螺旋变化总是偶数。据此线索，人们提出了很多可能的模式，其中一个就是所谓符号转化模式，这在第一章已经讲过了。

DNA 旋转酶有四个亚基，两个 α-亚基和两个 β-亚基。DNA 断裂后，两个 5′-P 末端共价连接于两个 α-亚基。这样，断裂双链 DNA 的两端连接于同一个酶分子上，而不致于造成开环；另一方面，它储藏了磷酸二酯键断裂所释放出来的能量。在转变符号后重新连接时不再需要活化。然而，这一过程仍要消耗 ATP 以促使酶构象的复原，这个酶的 β- 亚基可以结合 ATP 分子，具有 ATP 酶的性质。DNA 旋转酶的作用机制示于图 3-16。

这样，每当复制叉向前移动产生两个正超螺旋时，DNA 旋转酶作用一次产生两个负超螺旋。如果该酶能在复制叉前有一个结合位点，其作用应该是十分有效的，但是现在还不确切知道是否有这样的结合位点。

此外，在复制步骤的最后一步，如果是环形 DNA 的复制，必须有 DNA 旋转酶将环连分子(Catenane)分开，这在第一章中我们已经讲过。

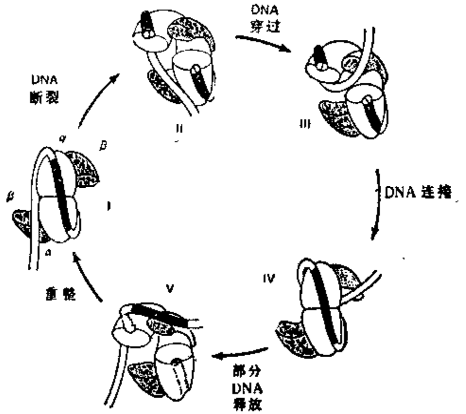



图 3-16 DNA 旋转酶引入负超螺旋的分子模型

I：DNA 双链以右手方向绕在四聚体的酶上 II：DNA 双链被切断，两个 α 亚基以其连接处为铰链转动，张开其靠近 DNA 断裂点的部分 III：未断的 DNA 双链穿过缺口 IV：连接断裂的 DNA；以左手方向绕在酶分子上 V：连接好的 DNA 释放出来。

## 第四节 DNA 复制的半不连续性[20~24]

### 一、DNA 半不连续性复制的发现

前面已经讲过，没有一种 DNA 聚合酶能按3′→5′方向使 DNA 链生长，也没有发现DNA 中含有三磷酸基团，更没有发现用 NMP 作前体来合成 DNA 的酶，因此链的生长方向总是5′→3′，而不可能相反。

链的生长按 5′→3′ 方向进行，又有两种可能性，一种是从亲本链的两端开始复制，各自有一个复制叉，当两个复制叉相遇时就能得到两个分子，每个分子有半截是单链状态(图 3-17)，实际上，细菌 DNA 在任何时刻只有大约 5% 可能处于单链状态，因而从两端同时复制是不可能的，而只能在同一个复制叉上以不同的方向复制新生链。

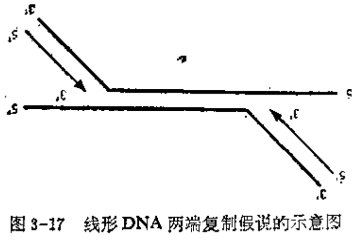


图 3-17 线形 DNA 两端复制假说的示意图

从一个复制叉上以不同的方向复制新生链又有两种可能，如图 3-18 所示。其中(a) 表示两条

新生链是分别复制的，一条是连续的，另一条不是连续的；而(b)表示模板转辙复制方式，其中新生链在复制叉处切断之前是相连的而且碱基是互补的，在变性后则可形成发夹结构。但实验证明，新生的 DNA 分子是完全可以变性的，因而只有 (a) 所示的情况才是可能的。

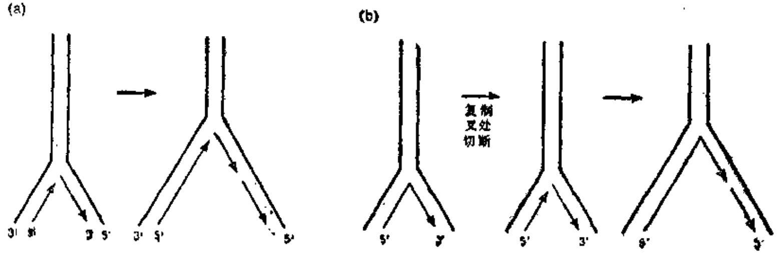



图 3-18 一个复制叉上复制的两种可能性

在图 3-18(a) 中所示的片断叫前体片断，1968 年冈崎发现了这些片断，因此又称为冈崎片断。冈崎用两种实验来证实这些片断的存在：一种是脉冲标记实验，一种是脉冲追踪实验。脉冲标记是以[$^3$H]dT 加入到 *E.coli* 培养基中，30 秒钟杀死细菌提取 DNA 进行碱离心分析。脉冲追踪实验是在脉冲标记的基础上(标记 30 秒钟)，将细菌立即转移到无放射性的 dT 中培养数分钟，然后提取 DNA 进行碱离心分析，其结果示于图 3-19。

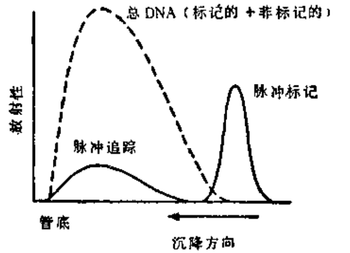



图 3-19 脉冲标记和脉冲追踪及其碱离心实验结果

结果表明，脉冲标记的片断大小为1000～2 000 个核苷酸。按图 3-18(a) 的复制模式，只应当有一半的新合成的 DNA 有冈崎片断，而图 3-19 中都是差不多大小的冈崎片断，这不能不令人吃惊。其实是，沿着复制叉前进方向合成的这条链(先导链)，其合成过程是连续的，在合成时，由于尿嘧啶的渗入，以致于实验时才产生冈崎片断。

上面在讲解核苷三磷酸前体时，曾经讲到，由于 DNA 聚合酶不能区别 dUTP 和 TTP (即 dTTP)，在 *E.coli* 体内有 dUTPase，可将 dUTP (一部分由 dUDP 磷酸化产生；一部分由 dCTP 脱氨氧化而产生)转变为 dUMP；然而 dUTPase 的作用并非能一个不漏地将 dUTP 转变为 dUMP，仍有极少数 dUTP 逃遁并渗入到 DNA 中去和A配对。(此外，细胞内 DNA 也以极缓慢的速度进行着 C 脱氨氧化成 U 的反应。但在新生的 DNA 链中，这可能不起主要作用)。这些在 DNA 中的 U 必须要去除，这就需要一系列修复体系的酶的共同作用，其中包括尿嘧啶 N-糖基酶(即尿嘧啶 N-糖苷酶，尿嘧啶 DNA 糖苷酶)，AP 内切核酸酶(即无碱基内切核酸酶，无嘌呤内切核酸酶)，外切核酸酶(polI)，DNA 聚合酶(polI)，DNA 连接酶等(具体反应机制见修复一章)。

由于体内尿嘧啶N-糖基酶作用较快，而 AP 内切核酸酶作用很慢，因此提取出来的没有尿嘧啶的磷酸二酯键很容易碱水解而断裂。在大肠杆菌细胞内，dUTP 与 dTTP 之比大约为 1:300，因此新合成 DNA 的链中平均大约 1 200 个碱基就可能有一个尿嘧啶存在。后随链也同

样发生尿嘧啶的渗入和切除事件。因而冈崎片断的大小实际上主要取决于U渗入的频率。于是，两条新生链在碱水解后产生差不多大小的冈崎片断。这种情况下的冈崎片断应该称为尿嘧啶片断，它并不能反映后随链合成的周期性起始的情况。

关于 dUTP 渗入 DNA，可以用突变菌株得到证明。缺少 dUTPase 的突变菌株($dut^-$)中，就有更多的尿嘧啶渗入 DNA，因而冈崎片断就小一些。缺少尿嘧啶 -N- 糖基酶的突变菌株($ung^-$)中，只有大约一半的新生 DNA 含有冈崎片断，而且这样的冈崎片断要大一些。新生的 DNA 链上虽有尿嘧啶的渗入，但由于缺少尿嘧啶-N- 糖基酶而不能去除尿嘧啶，也就使 DNA 样品在碱性离心时不致于在尿嘧啶处引起磷酸二酯键的断裂。

在 $ung^-$ 突变体中所产生的冈崎片断说明了在 DNA 复制过程中，一条新生链的合成是连续的，另一条新生链的合成则是不连续的。不连续的合成需要模板上的特别序列或其它信号，借以周期性地起始合成冈崎片断，然后再将这些片断连接起来。就总体来说，两条新生链的合成进度大体相同；但是在复制叉处，连续合成的链总是领先于不连续合成的链，因此前者称为先导链，后者称为后随链。因为不连续的合成的片断需要一个特别的信号以便进行周期性的起始作用，故有此差别。这可以用高分辨率的电子显微镜观察正在复制的 DNA 分子加以证实(图 3-20)。DNA 聚合酶Ⅲ全酶是一个具有双活性位点的非对称的聚集体，实现了先导链和后随链的同时复制。因此，二者的复制进度大体相同是理所当然的事。

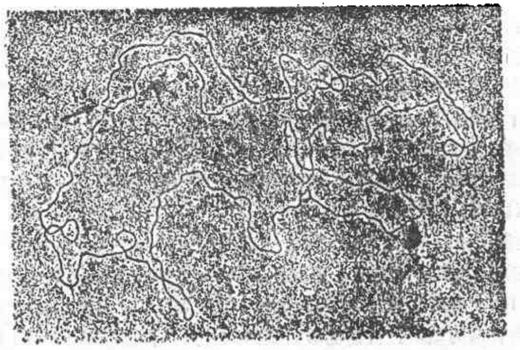


图 3-20　λ 噬菌体的 θ 复制的电子显微镜照片。箭头指处为单链部分

## 二、引物 (Primer)和引发酶 (Primase)

人们在确定各种 DNA 聚合酶所需的反应条件时出乎预料地发现，已知的任何一种 DNA 聚合酶都不能从头起始合成一条新的 DNA 链，而必须有一段引物。迄今为止，人们所发现的引物大多为一段 RNA，其长度和序列随着基因组的种类而异，在大多数情况下为 1～10 个核苷酸(表 3-3)。引物与典型的 RNA (如 mRNA)不同，它们在合成以后并不与模板分离，而是以氢键与模板结合。可见这种引物可能是由一种性质独特的 RNA 聚合酶所合成。这种合成 RNA 引物的酶就称为引发酶(注意："引物酶"是极其错误的名称)。*E.coli* 的引发酶是一条单肽链，分子量为 60KDa，每个细胞中有 50～100 个分子，它是由 dnaG 基因所编码的。该酶单

独存在时是相当不活泼的，只有在与有关蛋白质相互结合成为一个复合体时才有活性。这种复合体就称为引发体(见本章第六节)。

**表 3-3 后随链前体片断的RNA 引物**

| 基因组 | 引物长度 (nt) | 主要的引物的序列(中间 P 省略) |
|---|---|---|
| T7 | 1～5 | pppAC(C/A)$N_{1-3}$(N 主要为 A 和 C) |
| T4 | 1～5 | $pppACN_3$ |
| φ×174 | 1～5 (1nt 占 48%, 2nt 占 23%, 1～5nt 占 92%) | $pppAN_{0-4}$ |
| 海胆 | 1～大约 8 | (p)pp(A/G)$N_7$ |
| 多瘤病毒SV40 | 大约 10 | ppp(A/G)$N_9$ |
| 哺乳动物 | 大约 9 | |

在细菌中，先导链的引物和后随链中前体片断的引物在合成时需要不同的条件。实验表明，后随链前体片断引物的合成不受利福平(rifampicin)的抑制，而先导链引物的合成受到利福平的强烈抑制。人们已经知道，引发酶对利福平是不敏感的，而 RNA 聚合酶对利福平十分敏感。因此在很长时期中，人们认为后随链前体片断的引物是由引发酶合成的，而先导链的引物是由 RNA 聚合酶合成的。近来，Kornberg 等人认为，这是由于两种类型的引物合成是在不同的分子状态下进行的：先导链的引物合成需要引发双链 DNA，而后随链中前体片断引物的合成是在已经解链的状态下进行的。可能在大多数情况下，两种类型的引物都是由引发酶合成的，而 RNA 聚合酶的主要作用就是通过转录过程而解开局部的 DNA 双螺旋。这种作用就称为转录激活(transcriptional activation)。

总之，一个复制叉的诞生，关键是先导链合成的起始；先导链合成起始的关键又是引物的产生；引物产生的关键又是转录激活，也就是 RNA 聚合酶将复制原点的 DNA 序列转录成一系列长度不等的 RNA 短链。至于转录激活所产生的 RNA 能否作为引物来引发先导链的合成，至今尚未得出普遍的结论。目前能够肯定的只有大肠杆菌质粒 Col El DNA 复制时先导链的引物是由 RNA 聚合酶合成的。这类引物并不是整个的转录产物，而是经过 RNAaseH 部分降解后的片断。可能在大多数情况下，转录产物并不用作引物，其主要作用是打开复制原点处的 DNA 双螺旋以暴露出特定的 DNA 序列，然后引发酶和有关的蛋白质在此序列上装配成引发体，进而合成 RNA 引物。以上这两种引物产生方式示于图 3-21。

### 三、前体片断的连接

当 polⅢ 将一段 DNA 链延伸到前一个前体片断的引物 RNA时，它既不能继续合成 DNA，也不能水解 RNA 引物，而只是留下切刻或者缺口(究竟是切刻还是缺口或者随机是二者之一，都还未得到证实)。即使是留下切刻，由于RNA引物的 5′末端有三磷酸基团，没有一种连接酶能将 DNA 片断与 RNA 引物连接起来。这时，必须由 polI 的 5′→3′ 外切核酸酶活性来切除 RNA 引物(RNAaseH也参与这一切除过程)，同时 pol I 以其 5′→3′ 聚合酶活性将适当

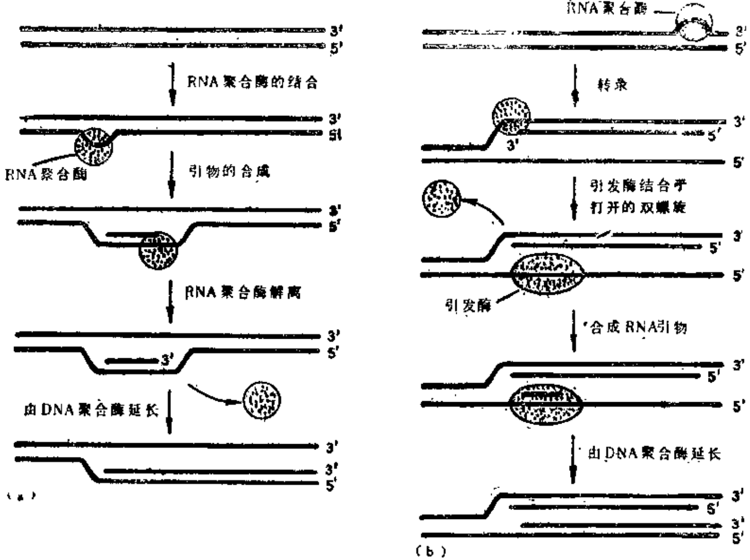



图 3-21 RNA聚合酶在引物产生过程中的作用

(a) RNA 聚合酶转录出RNA短链 经 RNAaseH 修饰(图中未表示出来)然后作为引物 (b) RNA 聚合酶转录出的 RNA 短链分开了 DNA 双螺旋使引发体得以形成，然后由引发酶合成引物

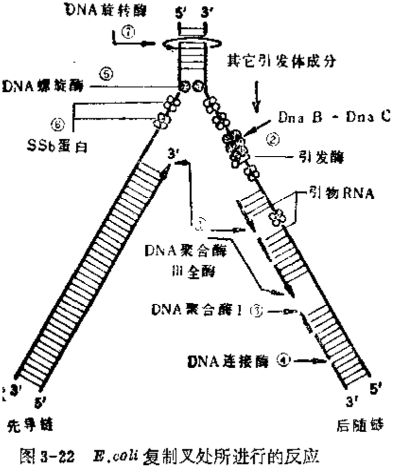



图 3-22 *E.coli* 复制叉处所进行的反应

① DNA 聚合酶Ⅲ全酶在两条新生链上进行聚合反应； ② DNA 引发酶(在引发体中)合成引物 RNA；③ DNA 聚合酶Ⅰ和RNAaseH切除引物RNA 并补上脱氧核苷酸； ④ DNA 连接酶将一个个冈崎片断连接到后随链上；⑤ DNA 螺旋酶解开双螺旋成为单链； ⑥ 单链 DNA 结合蛋白结合于解开的单链；⑦ DNA 旋转酶引入负超螺旋以解除复制叉前进所造成的超缠。

的脱氧核苷酸聚合上来，也就是说进行了连续的切刻平移或缺口充填。当 polI 前进到前一个片断的脱氧核苷酸部分时，有时也能将切刻再向前推进几个核苷酸，但由于 DNA 连接酶的竞争，很快封合了切刻。这个过程以及整个复制叉附近所进行的反应示于图 3-22。

为什么需要 RNA 作引物而不直接起始一条 DNA 链呢？这一问题仍然是一谜。我们已经知道，一切生物都要以极高的保真度复制其 DNA。从模板拷贝最初的几个核苷酸时，由于碱基堆集力很弱，其氢键结合能力也是很弱的，因而碱基对发生差错的机率就要大得很多，加上这几个核苷酸还没有与模板形成稳定的双链结构，DNA聚合酶的 3′→5′ 校对功能很难发挥作用。因而解决办法就是要用一种过渡形式，而这种过渡形式在不再需要时又容易被剔除。看来 RNA 是比较适合的，容易被 PolⅢ 识别而停止其聚合作用，又便于 PolⅠ进行切刻平移。切刻平移过程中发生错误的机率极少，因为 PolⅠ同样有 3′→5′ 校对功能。这个切刻最终为 DNA 连接酶所封合。以上所说，只不过是假说而已；不过，它也似乎是有道理的。

## 第五节　DNA 复制机构的复杂性[11～14,25～27]

由于 DNA 所携带的遗传信息关系到物种的根本特性，因此 DNA 复制过程所要求的精确度非常高。如果这一精确度不能得到保证，不但不能保持物种的根本特性，而且常常连生命也维持不了。因此，如果缺乏复制 DNA 的能力，对于任何一个细胞都是致命的。正因为 DNA 复制对于有机体是如此重要，以致于生物体演化出极为复杂的 DNA 复制机构。在大肠杆菌中，DNA 复制涉及到 30 种以上的基因产物(见表 3-4)。其中一些基因产物在一个特定的复制原点 *oriC* 序列上装配成复制体(将在下一节叙述)。许多与 DNA 复制直接相关的基因被称为 *dna* 基因；不过也有其他不同的名称。

由于 DNA 复制的重要性，有关 DNA 复制基因的突变体必然只能是条件致死突变，即该突变体在允许条件下 (如 37℃) 能进行 DNA 复制，而在非允许条件下(如 42℃) 则不能进行 DNA 复制。利用温度敏感突变体已鉴定出一系列的 *dna* 基因位点。根据这些突变体的表现行为可以分为两类：一类是缓慢停止突变体(slow-stop mutants，以 *s* 表示)，即该突变体在非允许温度下能完成正在进行的一轮复制，而不能开始下一轮复制。因此 *s* 突变体基因的产物必然是细胞从染色体复制原点起始一轮复制所必需的。另一类是迅速停止突变体 (fast-stop mutants，以 F 表示)，即该突变体在提高温度时则立即停止复制。因此这些突变体的基因的产物必然是 DNA 链的延长及冈崎片断的起始与连接所必需的。需要指出，某些 F 突变体基因的产物也是复制起始过程所必不可少的。尽管其表型是迅速停止复制。还有，用温度敏感突变体鉴定出来的 *dna* 基因并非都直接参与 DNA 复制，而是涉及到与 DNA 复制相关的过程。例如 *dnaF* 的产物后来证明是核糖核苷酸还原酶的一个亚基，现在称为 *rndA*；又如 *dnaS* 产物是 dUTPase，现在称为 *dut*。这两个基因都与 DNA 复制所需的前体供应有关。然而，尽管人们花了巨大的精力来分离 *dna* 突变体，但并未能鉴定出所有的 DNA 复制所需的蛋白质，例如 n′，n″，n，$\delta$，$\delta'$，$\chi$，$\psi$ 等以及一些目前尚不知晓的蛋白质。

为鉴定某个 *dna* 基因的作用，可以有两种方法，一种是将完全的复制体分离成单一组分，然后再重建；一种是从 *dna* 突变体中分离复制体系，然后再加入野生型的提取物观察复制活性的恢复情况。用这种体外互补试验可以鉴定各 *dna* 基因的产物。

表3-4 *E.coli*的复制基因

| 基　因 | 突变体的表型 | 功　能　或　基　因　的　产　物 |
|---|---|---|
| *dnaA* | S | *E.coli* 复制起始；分子量52KDa 聚集于 *oric* |
| *dnaB* | F,S | 链延长。引发体的组分，六聚体，分子量300 KDa，每细胞20个分子，具有螺旋酶活性 |
| *dnaC* | F,S | 复制起始和前体片断合成的起始。引发体组分；分子量25 KDa，六个单体结合于DnaB六聚体。 |
| *dnaD* | | 其实就是*dnaC*，符号*dnaD*不再使用。 |
| *dnaE* | F | pol Ⅲ 的核心蛋白；具有聚合和校对功能、分子量132 KDa，每细胞20个分子。 |
| *dnaF* | F | 核糖核苷酸还原酶，又称 *nrdA* |
| *dnaG* | F | 引发酶、合成前体片断的引物，单体 分子量60 KDa，每细胞75个分子 |
| *dnaH* | | 与*dnaA*类似，还未很好地鉴定。 |
| *dnaI* | S | 链的延长、细节不明。 |
| *dnaK* | S | 复制起始、细节不明、与果蝇的HSP70有相关性。 |
| *dnaN* | F | 链的延长，Pol Ⅲ 全酶的亚基 $\beta$、分子量37 KDa。 |
| *dnaP* | S | 复制起始、细节不明；PolⅢ 全酶的亚基。 |
| *dnaQ* | F | 复制的保真度；PolⅢ 的全酶的亚基 $\varepsilon$，分子量27KDa。 |
| *dnaZX* | F | 编码71 KDa的亚基 $\tau$，亚基 $\tau$ 能与模板结合；亚基 $\tau$ 在蛋白水解酶作用下生成52KDa的亚基 $\gamma$ 以构成 $\gamma$ 复合体。 |
| *dnaY* | F | 118bp，产物是稀有的 $tRNA^{Arg}$ |
| *polA* | F | 多功能的Pol I，链的延长；前体片断连接。 |
| *gyrA* | S,F | DNA旋转酶的 $\alpha$ 亚基、能引入负超螺旋，能消除正超螺旋，可为萘啶酮酸所抑制。 |
| *gyrB* | S,F | DNA旋转酶的 $\beta$ 亚基、能为香豆霉素所抑制。 |
| *lig* | F | DNA连接酶、封合切刻、连接前体片断，分子量75 KDa，每细胞300个分子。 |
| *dut* | | dUTPase |
| *rep* | F | 螺旋酶，在复制叉处解开DNA双螺旋，需水解ATP，分子量66KDa，每细胞50个分子。 |
| *ung* | | 尿嘧啶-N-糖基酶，从DNA上除去尿嘧啶。 |
| *rpoA,B,C D* | S | RNA聚合酶亚基，复制起始，某些体系中合成引物 |
| *hu* | | 刺激依赖 *oriC* 的DNA复制 |
| *ssb* | F | 单链DNA结合蛋白，能防止链间或链内形成氢键，复制叉的前进，引发过程，链的延长都不能缺少。四聚体，分子量74 KDa，每细胞有300个分子。 |
| *i(dnaT)* | F | 预引发体及引发体组分，1～2分子/复制叉，三聚体，分子量66 KDa，每细胞50个分子。 |
| *n'* | F | 预引发体及引发体组分，1分子/复制叉、分子量 55 KDa，每细胞有80个分子。 |
| *n''* | F | 预引发体及引发体组分。 |
| *n* | F | 预引发体及引发体组分。 |

注：突变体表型S表示条件突变体在非许可条件（温度）下复制缓慢停止。F表示条件突变体在非许可条件(温度)下复制迅速停止。

## 第六节　DNA复制的起始[11~14,28~53]

### 一、先导链合成的起始(上)：从新起始

关于从新起始合成DNA的机制，人们仍然知道得很少。迄今所知道的能普遍适用于细菌和噬菌体的一个规律是：从新起始合成DNA发生在特定的DNA序列即复制原点之中，原核

生物基因组一般只有一个复制原点，尽管有时主要复制原点因突变而丧失功能后会有次要的复制原点代替它；由 RNA 聚合酶所进行的转录激活是不可缺少的；不同的基因组还需要不同的蛋白质以识别其复制原点并形成引发体。下面，我们首先考察一下复制原点的特性、结构和功能，特别是 *E.coli* 的复制原点 *oriC* 和依赖于*oriC* DNA 复制起始的可能步骤；然后考察基本上依靠寄主复制机构，但也有自己的蛋白质因子参加复制的噬菌体λ的复制起始以及基本上依靠自己复制机构的噬菌体 T7 的复制起始。

1．复制原点与复制的起始，*oriC* 与 *E.coli* DNA 复制的起始　所有基因组DNA 的复制原点都处于双螺旋结构内部，就是线状 DNA 也不是从末端开始复制的(引物不是 RNA 的复制例外)。DNA 在复制原点处开始复制时，总是先出现一个复制泡。其中先导链已经引发并开始合成，与其模板形成双链结构，而另一条亲本链则被先导链置换出来。这样的由一条单链和一条双链组成的三元泡状结构，又叫置换噜噗或 D 噜噗(displacement loop)。只有先导链将另一条亲本链(即后随链的模板)的特别序列置换出来(以单链的形式)，才能产生后随链前体片断的预引发作用 (prepriming)。

最典型的 D 噜噗复制是哺乳动物线粒体的复制。在双螺旋的特定位点即重链的复制原点 $O_H$ 附近，以轻链为模板复制一条重链，而原来的亲本重链被置换出来成为 D 噜噗。在通常情况下，这段新合成的重链只有 500～600 b，有特定的 5′ 端，而 3′ 端则可能有 3～4 种不同的位置。这条新生的短链不稳定，很快被降解，但又很快重新合成，以经常保持线粒体的这一区段成为 D 噜噗。高等植物的叶绿体也有两个 D 噜噗。线粒体需要复制其 DNA 时，D 噜噗继续扩大；一直扩大到整个线粒体 DNA 的大约 67% 的位置时，被置换出来的亲本重链上的轻链复制原点 $O_L$ 才暴露出来，于是开始了轻链的合成。由于起步较晚，当重链合成完毕时，轻链才进展 30～40%。这时分离出两个环形 DNA 分子，一个是双链环，一个是不完整的双链环。各自完成复制合成以后再由 DNA 连接酶连接成封闭的双链环，进而由拓扑异构酶转变为超螺旋形式。这一过程示于图 3-23。

关于线粒体的复制机构及复制机制还不完全清楚。从线粒体的 D 噜噗复制可以看出，复制原点是这样一段 DNA 序列：它起始以一条亲本链为模板的 DNA 的合成；双向复制可能意味着每一条 DNA 链均有自己的复制原点；当两条链的复制原点非常靠近时即处于通常所说的一个复制原点之中时，才能表现出双向等速复制。由复制原点的性质决定了复制叉的数目和方向以及每个复制叉的进度。人们可能会天真地推想，在一个方向的复制叉中产生了先导链和后随链的前体片断，然后第一个前体片断沿着 5′→3′ 方向延长，就成为另一个相反方向的复制叉中的事实上的先导链。假若果真如此，则任何一个复制元 (replicon) 的复制都应表现为双向复制。而事实上，复制叉的数目、方向和进度表现出多种花样，甚至有完全为单向复制的情况，正如第二节中我们所看到的那样。单向复制以及其他多种事实导致了这样的建议：同一个 DNA 聚合酶 Ⅲ 复合体(复制体 replisome)同时复制了先导链和后随链(先导链和后随链同时复制模型将在下一节中叙述)；相反方向复制叉的启动，可能需要装配第二个复制体。

*oriC* 是细菌 DNA 复制调控的主要场所。DNA 复制速率的调节主要在于起始的频率，而 DNA 延长的速度则大体上是恒定的。还有证据表明，*oriC* 不但与复制起始有关，而且还与复制终止以及子染色体在子代细胞之间的分配有关。近年来，由于 *oriC* 的克隆与其核苷酸序列分析的进展和依赖于 *oriC* 的体外 DNA 合成起始系统的建立，使 *oriC* 的结构和功能的研究进入了一个新阶段。

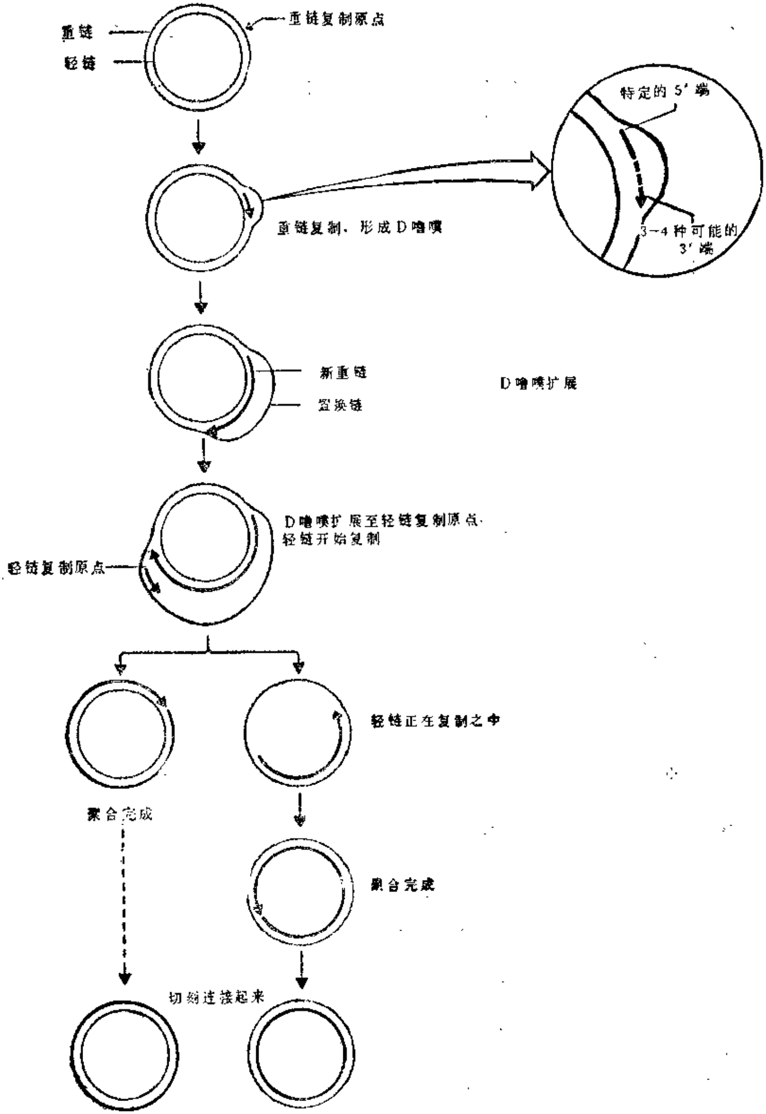



图 3-23　哺乳动物线粒体有一个稳定的D噜噗结构，每条DNA链有各自的复制起点，因而两条链的复制进度相差甚远。

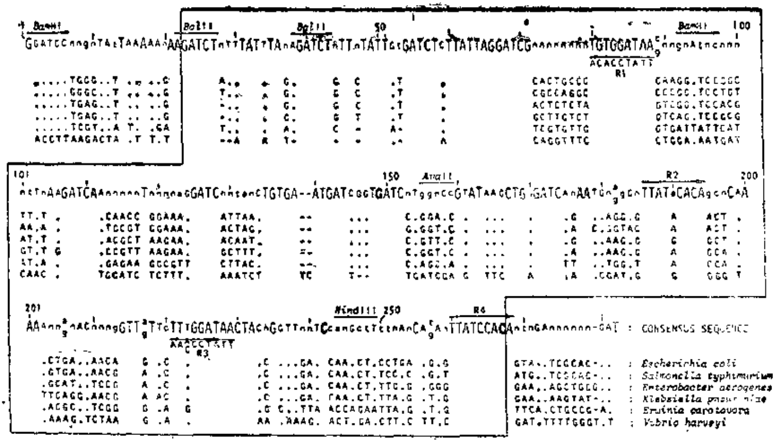


图 3-24 革兰氏阴性菌的最低限度复制原点的一致序列。大写字母表示六种细菌中全保守的核苷酸；小大写字母表示在五种细菌中都相同的核苷酸；小写字母表示六种细菌中有 3 或 4 种相同而且总共只有两种类型的核苷酸；n表示六种细菌中有三种以上的核苷酸类型或者虽只有两种核苷酸类型但有一种细菌在此处缺失。粗体字母表示单碱基替代即可导致 *oriC* 丧失功能的核苷酸。(·) 表示与一致序列相同，(-) 表示缺失。

通过缺失分析已经证明，*E.coli* 双向复制原点 *oriC* 至少包括 245 bp(+22～+267)，位于 *gidA* 基因和 16 KDa 基因之间，在 *E.coliK*12 株的遗传图上位于84分钟(图 3-24)。另外 5 种革兰氏阴性菌的 *oriC* 也已经确定下来；它们的序列与 *E.coli* 的 *oriC* 高度相关，可以看作是 *E.coli oriC* 的突变体，因为它们能够作为基因组的复制原点在 *E.coli* 细胞中起作用并使用*E.coli*的复制起始因子。其他 5 种 *oriC* 与*E.coli* 相比较，只有极少量的插入和缺失，而且集中在 +28～+53 这一范围内。在 6 种 *oriC* 中，全保守的序列高达 50%；6 种 *oriC* 序列中有 5 种相同的则高达 67%。

大肠杆菌型的 *oriC* 在结构上有如下特征：

(1) (A+T)含量占 56%。这一数值低于多数噬菌体复制原点中 (A+T) 的含量。

(2) 在 *oriC* 中，除了碱基序列高度保守之外，而在某些部分虽然碱基种类不十分保守，但碱基数目却是保守的。例如在 +92，+155，+244 附近可以允许碱基替代突变的存在，但不能允许碱基的插入和缺失。

(3) 在 *oriC* 中，有 9～14 个 GATC 位点(大肠杆菌有 11 个)，其中有 8 个 GATC 位点在 6 种 *oriC* 中的位置完全保守。实验表明，*oriC* 质粒对于 *dam*⁻ 菌株的转化率极低，而且从 *dam*⁻. *E.coli* 菌株中分离出来的 *oriC*质粒在依赖于 *oriC*的体外起始复制体系中不能很好地起始复制。因此，GATC 位点中腺嘌呤的甲基化对于 *oriC* 的功能是不可缺少的。

(4) 有四个高度保守的 DnaA 识别位点 R1，R2，R3，R4，其中 R1 和 R4 互为完全反向的重复序列，而且在 6 种 *oriC* 中的序列完全相同。R3是 R1 的正向重复。R2 是 R4 的正向重复。另外 R5 在 5 种 *oriC* 中也是高度保守的。这些位点是 DnaA 与 *oriC* 的结合所不可缺

少的。

(5) 在 *oriC* 中没有大于 20 个密码子的开放阅读框。

依赖于 *oriC* 的体外复制起始系统必须要有 DnaA 的存在，而且对利福平是敏感的。这就表明 DnaA 和 RNA 聚合酶是不可缺少的。根据 A.Kornberg 等人的研究，*oriC* 的复制起始还需要 DnaB，DnaC，DNA 引发酶，DNA 旋转酶，DNA 聚合酶Ⅲ全酶以及 HU 蛋白，拓扑异构酶 I，SSB 蛋白，RNAaseH 等辅助成分。综合近年来的研究结果，*E.coli* DNA 的复制起始可能包括如下步骤：

(1) 转录激活：由 RNA 聚合酶发动的转录作用主要从 16 KDa 基因的启动子 (P2) 以及 *asnC* 基因的启动子 (P1) 开始，由右至左进入 *oriC*。这些转录物在 *oriC* 内大约有 20 个终止点簇，其中大部分位于 *oriC* 的左半边。在 *oriC* 左半边的终止位点与 RNA-DNA 转变位点十分吻合。当然这并不能说明引物就是 RNA 聚合酶的转录物，上一节中所述的两种可能性均不能排除。这样大肠杆菌 DNA 复制时首先启动的是反时针方向 (左向)复制叉。在 *oriC* 区域，并未发现顺时针方向复制叉的 RNA-DNA 转变位点。据推测，有可能在左向复制叉的第一个冈崎片断上形成右向复制叉的复制体。另一种可能性是右向复制叉的先导链是在 *oriC* 以外引发的。

(2) 起始复合物的形成，需要负超螺旋的 *oriC*，ATP，DnaA，HU 蛋白，TopI。DnaA 结合于R1-R4位点；HU蛋白识别并刺激 *oriC* 复制而抑制其他潜在的复制原点上的复制；TopI 也是 *oriC* 特异性复制所必需。在这一阶段，ATP 是必需的，但不被水解，因为用不可水解的 ATP同系物 ATP-$\gamma$S 代替 ATP 同样有效。

(3) 预引发体的形成：这一步骤的前提条件是 DnaA 已结合于 R1-R4 位点。DnaB-DnaC 六聚体与 *oriC* 结合成预引发体。

(4) RFI* 的形成：在 DnaB 的螺旋酶活性和 DNA 旋转酶的拓扑异构酶活性作用下，以水解 ATP 为能源，促进大范围的解链 (有 SSB 蛋白与单链 DNA 结合)，在 *oriC* 区域负超螺旋提高 10～15 倍，形成一个具很高迁移率的复制形式，称为 RFI*。

(5) 引发体的形成：高度解链的模板与蛋白质的复合体促进 DNA 引发酶加入进来形成引发体，然后合成引物 RNA。

(6) DNA 的合成：这项任务由DNA聚合酶Ⅲ全酶担任，其中 $\alpha$-亚基为聚合酶，$\varepsilon$-亚基为 $3'\rightarrow5'$外切核酸酶，$\beta$-亚基保证了全酶作用的进行性(processivity)，而 $\tau$-亚基和$\gamma$-复合体则保证了 $\beta$-亚基作用的发挥。

2．噬菌体 λ DNA复制的起始　噬菌体 λ 的 DNA 复制在早期阶段和晚期阶段采取不同的方式，在早期为 $\theta$ 复制，在晚期为滚环式复制。我们这里只讲其早期复制。λ 的复制原点处的结构示于图 3-25 中。λ 的复制原点包含三个部分：四个高度保守的 19 个 bp 组成的正向重复序列，称为 *ori* 重复序列；约 40 bp 组成的富含 A/T 序列，其中 A/T 占 80%；28 bp 的回文序列，中间 4 bp为不对称部分。

现已知道，由 $p_{R}o_{R}$ 发动的转录是 λ DNA 复制不可缺少的步骤。如果 λ 阻遏蛋白结合于 $p_{R}o_{R}$，则 λ 的复制即被阻断。转录作用除了使复制原点解链外，还转录了 λ 复制所必需的基因 O 和基因 P。基因 O 和基因 P 互相重迭 4 bp。基因 O 产物 298 个氨基酸，分子量 33 700；基因 P 产物 233 个氨基酸，分子量 26 500。从 $p_{R}o_{R}$ 起始的转录为本底水平的转录。O 蛋白非常不稳定，很容易被水解，而 P 蛋白则很稳定。O 蛋白的结合位点是四个 *ori* 重复序列。当 O 蛋

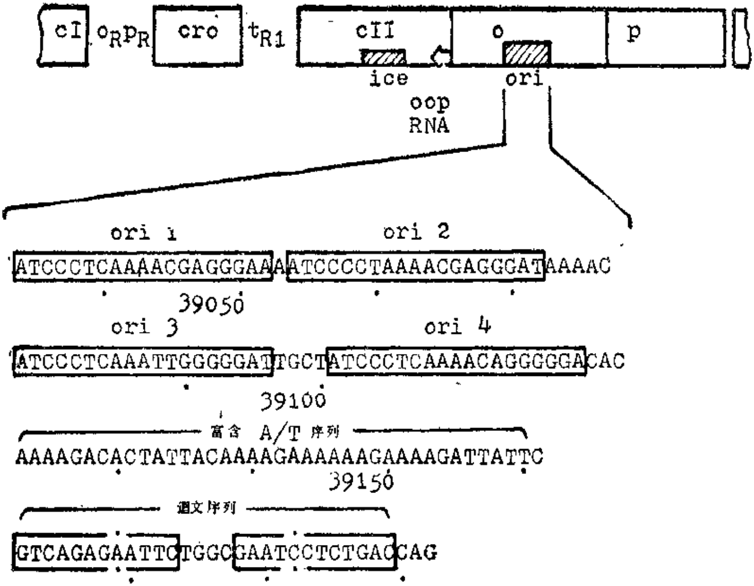



图 3-25 噬菌体 λ 的复制原点的结构，包括四个复制原点重复序列 (*ori*1～4)，富含 A/T 序列以及迴文序列

白含量较低时，优先结合于重复序列 2 和 3。DNA 的正常复制要求四个重复序列都必需被 O 蛋白饱和。遗传学实验表明，O 蛋白与 P 蛋白可以相互作用，体外实验尚未证实这一点。P 蛋白与 *dnaB* 基因产物可以相互作用。对于 DNA 复制起始所必需的序列大约有 350bp，它包括全部四个 *ori* 重复序列和富含 A/T 序列。而富含 A/T 序列后面的迴文序列以及前面的 *ice* 和 *oop* 序列(人们曾建议过它们在 DNA 复制中的作用)并非为 DNA 复制所必需。当然，这并没有否定这些序列在体内复制可能会有某种增进或控制的功能。用 [$\alpha$-$^{32}$p] dCTP和 ddCTP 试验表明，引物 RNA 在紧靠四个 *ori* 重复序列的左边转变为 DNA (同位素渗入量最高) 而且是双向复制的(同位素放射性的双向梯度下降)。

3. 噬菌体 T7 DNA 复制的起始　T7 是一种中等大小的噬菌体，是由 39 936bp 组成的线状双螺旋 DNA，含有差不多等量的四种碱基，没有异常碱基，有 160bp 的末端冗余 (terminal redundency)。T7DNA 的复制可以分为三个阶段：①复制起始；②复制叉的移动；③连环分子的形成和处理。这里我们重点讲解复制的起始问题。

正在复制的 T7 DNA 置于电子显微镜下观察时，其复制原点似乎在离末端 17% 的地方，而实际上是在 14.51～14.91 之间。只有了解 T 7 复制的起始机制和复制叉的移动情况才能解释这一差异。T7 复制原点位于基因 1 和基因 1.1 之间，其中包括基因 1.1 的两个启动子和一段 61 bp 的富含 AT 的序列(图 3-26)。基因 1.1 两个启动子各为 23 bp 高度保守的序列。从启动子 1.1 B 到核糖体的结合位点(图中标以 rbs)之间共有 61 bp 的富含 AT 的序列 (A+T 占

78%)，其中含有七个迴文序列 TTAA。此外，还有一个基因 4 产物的识别位点 3′ CTGGG5′。然而基因 4 产物是否在此合成引物，现在还不能肯定。噬菌体 T7 DNA 的复制起始，必需的酶有三个：T7RNA 聚合酶，T7 DNA聚合酶，基因 4 蛋白。下面分别叙述这三种酶的性质。

T7 RNA 聚合酶是基因 1 的产物，分子量 98 100，为一条单肽链，其启动子为一种 23 bP 的高度保守的启动子序列，T 7 共有 17 个这样的启动子，唯有靠近复制原点的基因 1.1 有两个这样的启动子。T7 RNA 聚合酶是基因 1.1 直到最末尾的基因转录所不可缺少的。

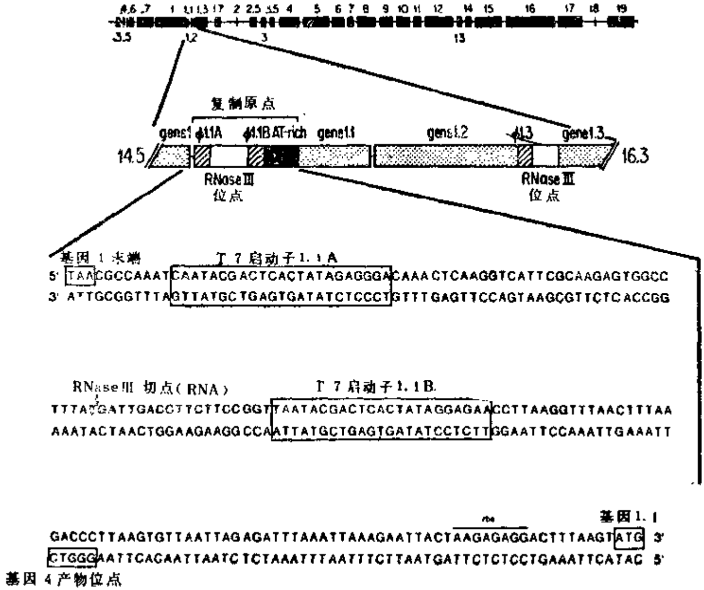



图 3-26　噬菌体 T7 的 DNA 复制原点

T7 DNA 聚合酶由两个亚基组成。一个是 T7 的基因 5 蛋白，分子量 84KDa，单独存在时除了具有单链 DNA 的 3′→5′ 外切核酸酶活性外，还有进行性很低的聚合酶活性，在催化 1～50 个脱氧核苷酸的渗入之后即与模板脱离。另一个亚基是噬菌体诱导出来的寄主基因 trxA 编码的硫氧还蛋白(thioredoxin)，分子量为 12KDa。两个亚基一经结合(按 1:1 比例)，则成为具有很高进行性的 DNA 聚合酶，还具有单链、双链 DNA 的3′→5′ 外切酶活性。

基因 4 蛋白有两种分子形式，一种分子量为 58 KDa，另一种分子量为 66 KDa，各自从基因 4 的 mRNA 上相距 189 b 的两个起点翻译而成。基因 4 产物是多功能的酶：一是依赖于单链 DNA 的核苷 -5′- 三磷酸酯酶活性 (dTTP 酶活性最高)，二是螺旋酶活性，三是引发酶活性。两种形式的基因 4 蛋白具有不完全相同的酶学性质。低分子量形式具有核苷-5′-三磷酸酯酶活性并且具有进行性的螺旋酶活性，而其引发酶活性极低。高分子量形式具有很高的引发酶活性，这种酶活性不是进行性的，而是分散性的(distributive)。高分子量形式的引发酶活

性在很大的范围内又与低分子量形式的含量成正比。这就表明，两种分子形式结合在一起才能具有引发酶的功能。可能是低分子量形式与模板结合，高分子量形式再与低分子量形式结合。当引物合成以后，高分子量形式即解离下来。然后高分子量形式与低分子量形式再度结合。在这反复的结合过程中，需要 T7 基因 2.5 的产物，即 T 7 的单链 DNA 结合蛋白。

缺失突变的研究表明。富含 AT 区域对于 T7DNA 复制起始是必不可少的。基因 1.1 的两个启动子至少得保留一个。根据现有资料，噬菌体 T7 在复制原点处 DNA 复制的起始可以用图 3-27 来描述。

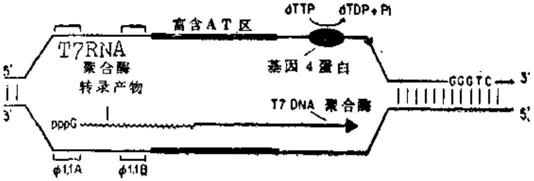



图 3-27　在复制原点 T7DNA 复制的起始机制(这里假定引物就是 T7RNA 聚合酶从启动子 1.1A 开始转录的 RNA)

用纯化的 T 7DNA 聚合酶进行的试验表明，T7 RNA 聚合酶可以利用基因 1.1 的两个启动子中的任何一个发动转录。这个转录物延伸到富含 A/T 区域的某个位点时，就被 T7 DNA 聚合酶用作引物以合成DNA，这条 DNA 链即为先导链。由合成 RNA 到合成 DNA 的转变过程中，富含 A/T 的 DNA 序列起重要的作用。由于转录作用，置换出一条亲本 DNA 链，基因4 蛋白就马上与之结合并水解 dTTP 等核苷三磷酸成为核苷二磷酸和无机磷酸借以推动基因 4 蛋白从 5′ 端向 3′ 端移动。由于基因 4 蛋白的螺旋酶活性，解开 T7DNA 双螺旋，使T7 DNA 聚合酶能够继续向前进行复制，当基因 4 蛋白遇到 5′GGGTC3′ 和 5′TGGTC3′ 的序列时，则行使其引发酶功能，在此处合成互补的引物 pppACCC 或 pppACCA，然后由 T7 DNA 聚合酶进行后随链前体片断的合成。左向复制叉启动的机制还不清楚，可能需要右向复制叉中第一个冈崎片断延伸至富含 AT 区域时，才能装配出左向复制叉的复制体。然而右向复制叉中第一个冈崎片断产生需要相当的时间。因而左向复制叉的启动要落后不少时间。这样，左右两个复制叉的中点位置必然出现在真正原点的右边，这也就是电子显微镜下所见到的 17% 的地方。

从以上 *oriC*，噬菌体 λ 和噬菌体 T7 DNA 复制的起始可以看到，先导链合成的重新起始，是染色体复制周期的开始，必然是处于严格控制之下的。参与这种控制的因素有两个方面：一方面是复制原点的特异性结构，另一方面是识别这种结构的特异性的蛋白质因子。例如 *oriC* 的 R1-R4 位点和 DnaA，噬菌体λ的四个 *oriC* 重复序列与 O 蛋白质，噬菌体 T 7 的 *ori* 区域与基因 2.5 蛋白。转录激活是必须的，它不仅解开了复制原点的双螺旋，而且有利于引物的合成或直接提供引物；有时还转录了控制复制起始的某些蛋白质因子的基因。复制原点除了含有特异的可为蛋白质因子识别的序列之外，富含A/T序列几乎是必不可少的。其原因可能是：(1)富含 A/T 区段呼吸作用强烈，DNA 易于解链；(2)在富含 A/T 的序列中，容易产生 RNA →DNA 的合成转变。

## 二、先导链合成的起始(下)：共价延伸

关于共价延伸，前面我们已经讲到滚环复制。这是目前关于先导链合成起始机制中研究得最深入的一个。由于它是在一条断裂的亲本链的 3′-OH 上不断地发生 DNA 聚合作用，因而叫共价延伸。由于断裂的 3′-OH 端仍然与其互补链形成双螺旋结构，因而没有合成 RNA 引物的必要。合成先导链的模板总是呈环状存在于分子中。

有一种滚环复制模型叫做噜噗滚环复制 (looped rolling circle replication)。它用于双链环状分子复制单链环状分子。噬菌体 $\phi\times174$ 就是利用这种方式产生其后代的。噬菌体的基因 A 蛋白首先结合于负超螺旋形式的 FRI DNA 的复制原点 (4299～4328)上的结合位点（又称为钥匙位点 key sequence)。由于负超螺旋和富含 A/T 序列的存在，A 蛋白的结合诱导了其识别位点及富含 A/T 序列的熔解。然后 A 蛋白有可能转移到其作用位点，在正链的第 4305 和第 4306 个核苷酸之间切断磷酸二酯键；与此同时，A 蛋白的酪氨酸羟基与断裂的 5′ 端核苷酸(A)形成磷酸酯键而共价相连。断裂链的另一端为自由 3′-OH 末端(图 3-28)

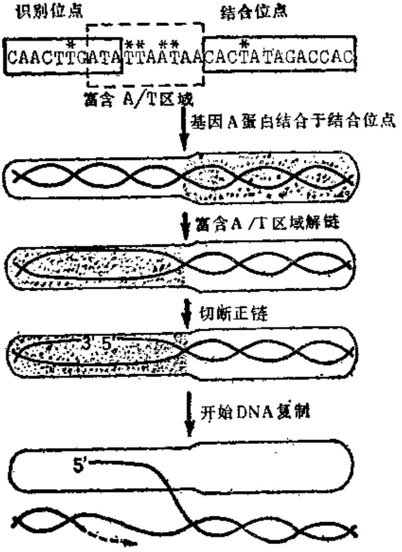



图 3-28　$\phi\times174$ 的基因 A 蛋白与复制原点(4299-4328)的相互作用模型，带星号字母表示非保守序列。

由于 Rep 蛋白，ssb 蛋白和 PolⅢ 的协同作用，(+)链的 3′-OH 不断延伸，而带有 5′-P～A 蛋白的一端不断被置换，ssb 蛋白不断地结合到甩出的单链上并改变其构型，趋向环化，直到基因 A 蛋白识别复制原点序列，将链切断，并利用上次水解磷酸二酯键所得到的能量将这一段 DNA 环化，然后基因A蛋白又重新连接于双链环（+）链的 5′-P 末端以进行下一个循环(图 3-29)。要注意噜噗滚环复制与一般的 σ 复制不同。被置换的链决不会超过原有环的长度，而且是恰好等于该环的长度，其原因就在于基因 A 蛋白对于单链状态的复制原点序列的有效识别。

在 *E.coli* 的 Hfr 雄性细胞和雌性细胞的长时间的接合过程中，常可以发现某些基因第二次进入雌性细胞的现象，这也就是基于 σ 复制所产生的连环分子 ABC……XYZAB。

在真核生物中发现的基因扩增现象，其原理也是滚环复制。例如非洲爪蟾的卵母细胞的前体细胞内有 500 个 rRNA 基因单元，通过大约 72 小时的扩增，变为 $2\times10^6$ 个基因单元，扩大 4000 倍。这时，rDNA 占整个细胞核 DNA 含量的 75%，因而使卵母细胞能装配 $10^{12}$ 个核糖体，以适应受精卵迅速分裂所需要的大量蛋白质的合成。

在扩增之前，rDNA 区段由 500 个rDNA 单元以前后相连的方式组成，每个单元包括一个

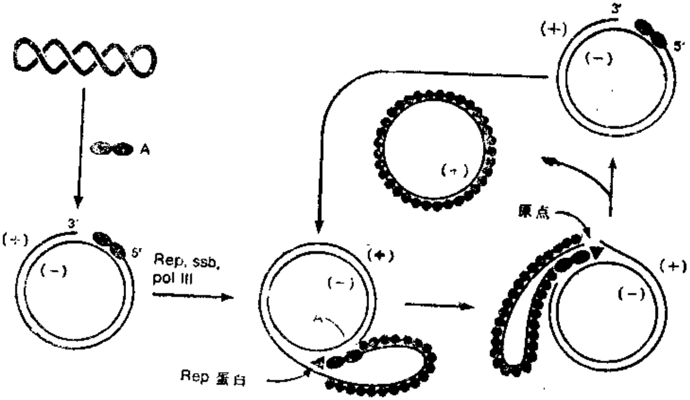



图 3-29 $\phi \times 174$ 的噬菌滚环复制

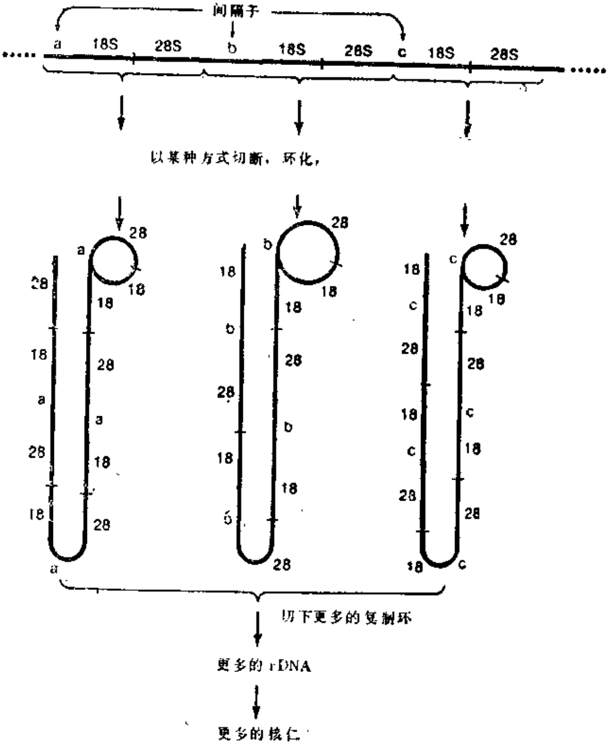



图 3-30 rDNA 的扩增示意图

18S RNA基因和一个28S RNA基因，每个基因之间都有长短不一的间隔子(spacer)(图3-30上)。整个rDNA区段包括在唯一的核仁中；而在扩增后，细胞中出现了数百个核仁。

在基因扩增以后，在细胞中发现许多小的环形DNA和正在复制的$\sigma$形复制环。奇怪的是这些环中可以包含2～4个rDNA单元，甚至多到16个；而且在18S RNA基因和28S RNA基因之间的间隔子通常只有一种大小。由此可以推测，由染色体上切下来一个rDNA单元然后环化(很可能切的同时就环化)通过滚环复制产生连环分子；再从这些连环分子上再切下一个单元或几个单元作为复制环再滚环复制。因而产生只有单一间隔子的多单元产物。从扩增因子的计算上也支持这种设想。因为一个rDNA单元长4.3 μm，要通过滚环复制形成一个单位大小的链大约需经半小时，而基因扩增大约72小时，这样扩增因子只能为144，况且实际上不可能所有的rDNA单元都从染色体上切下成为复制环。如果只有从染色体上产生复制环，那末扩增因子比144还要小，然而实际上的扩增因子是4 000，因此必然存在许多扩大再生产的"子工厂"。这样的基因扩增过程可以用图3-30来表示。

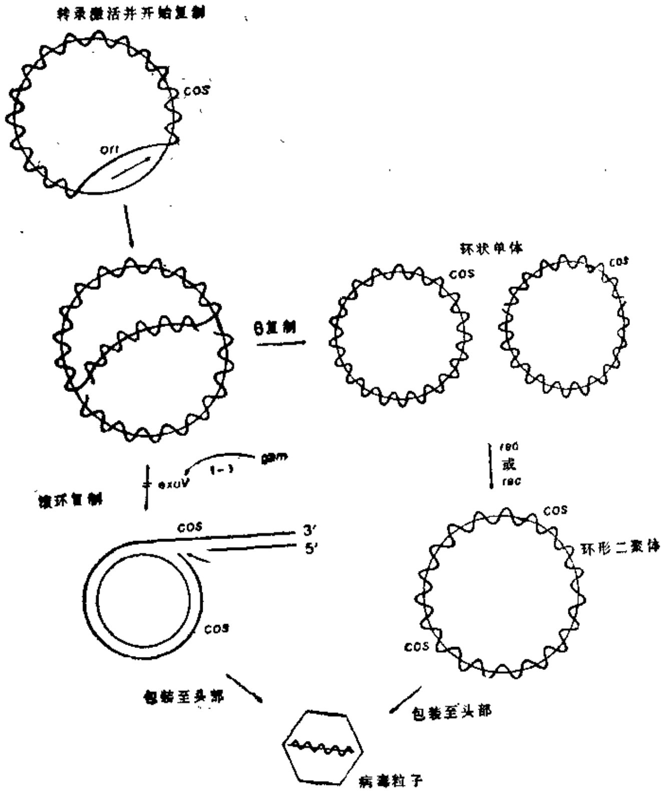



图3-31　λ噬菌体的两种复制方式

λ 噬菌体在其裂解生长的后期常采用滚环复制来产生其后代噬菌体线形DNA。*E.coli*的ExoV(RecBCD) 能阻碍这一过程；但 λ 的 *gam* 基因产物能使 ExoV 失活。滚环复制产生线形的连环体分子 (concatamers)，然后由 λ 的内切酶识别其 cos 位点产生线性 λ 分子包装到蛋白外壳中。θ 复制所产生的环状单体 (负超螺旋) 不能直接进行包装。必须通过重组作用先形成环状二聚体再由内切酶在 cos 位点切出线形分子然后再进行包装(图 3-31)。

### 三、后随链的前体片断的起始

关于 *E.coli* 的后随链前体片断的起始机制，人们的研究还不够系统，原因在于复制体系十分复杂。在 80 年代中期之前，人们把注意力主要集中在若干单链 DNA 的 *E.coli* 噬菌体上。这些噬菌体复制的第一步就是单链环状 DNA 转变为双链环状 DNA (RFⅡ)，进而成为双链环状超螺旋 DNA(RFI)。这些噬菌体在进行复制的过程中，大部或全部使用寄主的复制酶系和蛋白质。因而研究这些噬菌体 DNA 的复制对于了解 *E.coli* DNA 复制机制应该是十分有用的。研究得较多的有 M 13、G4、$\phi\times174$ 等噬菌体。

M13 DNA 有6 407 个碱基。从其第 5 480 至第 5 820 碱基之间有五个发夹结构，其中以第三个发夹结构最为重要(图 3-32)。这一发夹环有 59 个碱基。M 13 噬菌体在进入寄主细胞后，寄主的单链DNA 结合蛋白(SSB 蛋白)就结合于单链部分。而 RNA 聚合酶能识别发夹结构，从第 5 756 碱基开始转录出 20～30 碱基的 RNA，从而破坏了发夹结构，于是转录亦告停止。SSB 再结合于暴露出来的单链部分。然后，DNA 聚合酶Ⅲ以转录出的 RNA 为引物，开

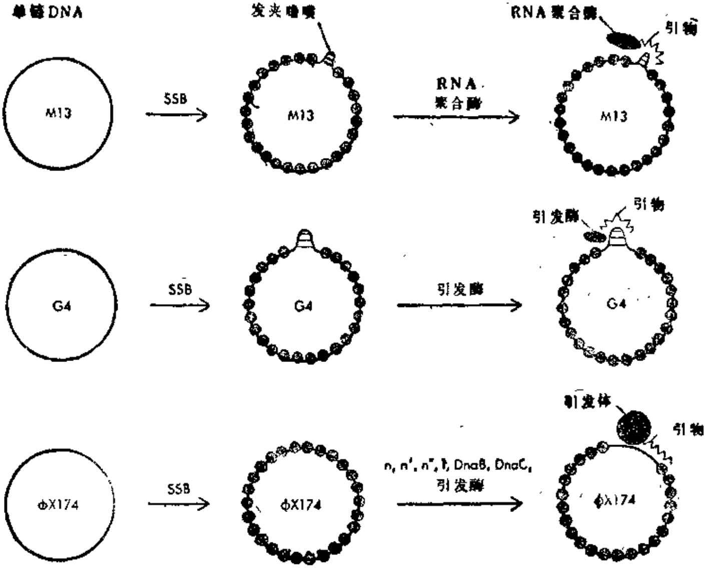



图 3-32 在三种单链 DNA 噬菌体的负链复制过程中，使用不同的机制合成其引物 RNA。

始互补 DNA 链(负链)的合成。由于引物 RNA 是由寄主的 RNA 聚合酶合成的,因而M13的互补的负链的合成对于利福霉素是敏感的。

G4 的单链环形 DNA 有 5 577 个碱基。在其基因 F 和 G 之间有一段与负链复制有关的序列。该序列两端为两个同向重复序列，中间包含三个发夹结构。G4 的负链合成的起始与 M13 很相似,唯一的差别是引物RNA是由寄主的引发酶(*dnaG* 基因所编码)合成的(图 3-32)。这样,G4 负链合成的起始不受利福霉素的影响。引发酶合成的引物 RNA 长短不一，可以长达 29 个碱基,少则只有 6 个碱基。这可能主要取决于 DNA 合成系统是否跟得上。

从 M 13 和 G 4 的负链复制起始机制来看,它们的复制原点处都有发夹结构，特别是对于这些单链环 DNA,发夹结构可能更为重要,也可能由于其发夹结构的存在，它们的复制机构较为简单，基本的要素为 SSB 蛋白、引发酶(或 RNA 聚合酶)和 DNA 聚合酶Ⅲ以及 DNA 聚合酶 I。然而,单链环状 $\phi\times174$ DNA 的复制机构要复杂得多，也可以说是迄今人们研究过的最为复杂体系。由于其基因中没有一个与负链复制有关,复制所用的酶和蛋白质全部来自寄主，因而在较长时间内人们认为 $\phi\times174$ 负链的复制机制应与其寄主大肠杆菌非常相似。然而近年来的研究表明,*oriC* 复制的引发并不需要蛋白质 n,n′, n″。

$\phi\times174$ 的环状单链 DAN 有 5 386 个碱基。当它进入寄主细胞后，其单链 DNA 上全部结合有 SSB 蛋白。$\phi\times174$ 的复制原点像 G4 一样，也在基因 F 和 G 之间。约 55 个碱基，有一小段发夹结构，n′ 在 DnaB 和 DnaC 存在时能识别这个序列。在这个序列上，n′、DnaB、DnaC、i、n、n″ 装配成为一个复合体,这个复合体就叫做预引发体(preprimosome)。在这个预引发体中,DnaB 为六聚体,每个单体又各结合一个 DnaC 分子。当 ATP 存在时，DnaB 与单链 DNA 有很高的亲和力。如果 ATP 水解成 ADP 和Pi ,则 DnaB 与单链 DNA 的亲和力下降。当预引发体在 55bp 处形成之后,引发酶加入进来成为引发体(primosome)。由于引发体中 DnaB 能改变局部 DNA 结构,造成引发酶合成引物 RNA 的条件。在引物合成以后，n′ 水解了与之结合的 ATP,使 DnaB 发生变构，降低了与单链 DNA 的亲和力，使整个引发体(或许至少DnaB,引发酶,n′三个组分),在单链 DNA 上沿着 5′→3′ 方向移动,直到发现其引发位点再合成引物 RNA。在移动过程中,SSB 蛋白被不断置换，过后又重新结合上去。是什么组分发现引发位点呢？体外实验表明。当 $\phi\times174$ DNA没有 SSB 蛋白结合时,引发酶只要在 DnaB 存在时就能识别各个引发位点并合成引物。这表明,引发位点是由 DnaB,引发酶复合物识别的。当 $\phi\times174$DNA 中加入 SSB 时,这一识别能力就丧失了。这可能是因为 DnaB，引发酶并不具备在 DNA 链上移动并置换SSB的能力。在大肠杆菌细胞内，任何一个单链 DNA 都会有 SSB 结合。因此，细胞内的 DNA 复制一直是在 SSB 存在下进行的。在 $\phi\times174$ 负链复制过程中,引发体移动方向与负链合成方向相反。

在 $\phi\times174$ 中,预引发位点只有一个,而引发位点有许多个。而在 M13、G4 中，合成一个引物,然后 DNA 合成到底。在负链复制的其余步骤上，这三个大肠杆菌噬菌体完全一样，由 DNA 聚合酶Ⅲ以引物 RNA 为依托,合成 DNA 直至引物 RNA 处；由 DNA 聚合酶 I 的5′→3′ 外切核酸酶活性水解引物 RNA 并代之以 DNA，最后由DNA连接酶连接成为共价封闭环。再继而由 DNA 旋转酶形成负超螺旋。这时有负超螺旋的双链共价封闭环即为这三个单链环状噬菌体的复制形式 I (RFI)。在连接酶连接之前，带有切刻的双链环称为复制形式 Ⅱ (RF Ⅱ)。

$\phi\times174$ 的负链复制的引发体的装配模型示于图 3-33。

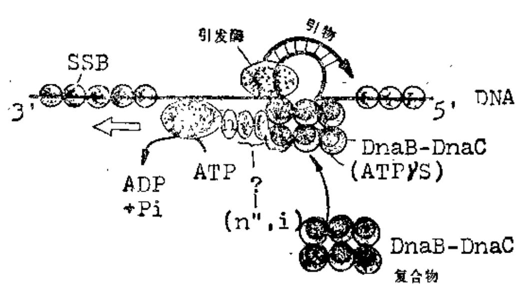



图 3-33 φ×174 负链复制中引发体的装配模型

φ×174DNA的复制，主要可以分为三个阶段：

1. 亲本单链 DNA(+)链转变成共价闭合环状双链超螺旋DNA(RFI) 这一阶段通常写作 SS→RF。这一阶段的反应发生在 RF1 转录之前，所以这一段复制过程全部使用寄主的蛋白质(包括酶)。这一过程在上面已经讲过。

2. 多拷贝 RFI 的合成 这一段的特点在于多功能的基因A蛋白从 RF1 的(-)链转录。注意 φ×174 的(-)链是唯一能够转录的链。在复制过程所使用的蛋白质中，只有A 蛋白是噬菌体编码的，这一阶段的具体过程在上一节已经讲过。

3. 后代单链 DNA (+)链的合成 这一阶段通常写作 RF→SS。从真正的 DNA 复制意义上来说，这一阶段的 DNA 复制是第二阶段的一部分，其实主要是噬菌体颗粒的装配。从嗜噗滚环复制出来的 (+)链 DNA 到底是用于"扩大再生产"还是直接用于颗粒的装配，二者之间并无一道开关，而是受装配系统蛋白质浓度调节的，当噜噗滚环达到每细胞 35 个时，才出现较多的装配系统的蛋白质。捕获后代(+)链进入噬菌体头部。

现将 φ×174 的全部复制周期列于表 3-5，并绘成图 3-34。

表 3-5 φ×174 的复制周期

| 阶 段 | 时间(分,在 33℃) | 变 化 |
|---|---|---|
| SS→RF | 0～1 | 吸附于细菌表面，穿透细胞壁。病毒单链转变为亲本复制形式(RF1)，RF1 转录。 |
| RF→RF | 1～20 | 亲本 RF1复制，产生 60 个后代 RF1 |
| | 25 | RF1 继续复制，寄主 DNA 合成停止 |
| RF→SS | 20～30 | 从后代 RF1 形成大约 35 个滚环，然后滚出大约 500 个单链分子；这些单链 DNA 被装入噬菌体颗粒中 |
| | 40 | 发生细胞溶解 |

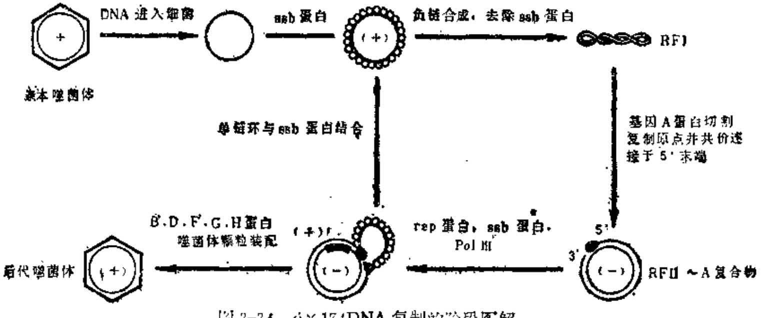



图 3-34 φ×174DNA 复制的阶段图解

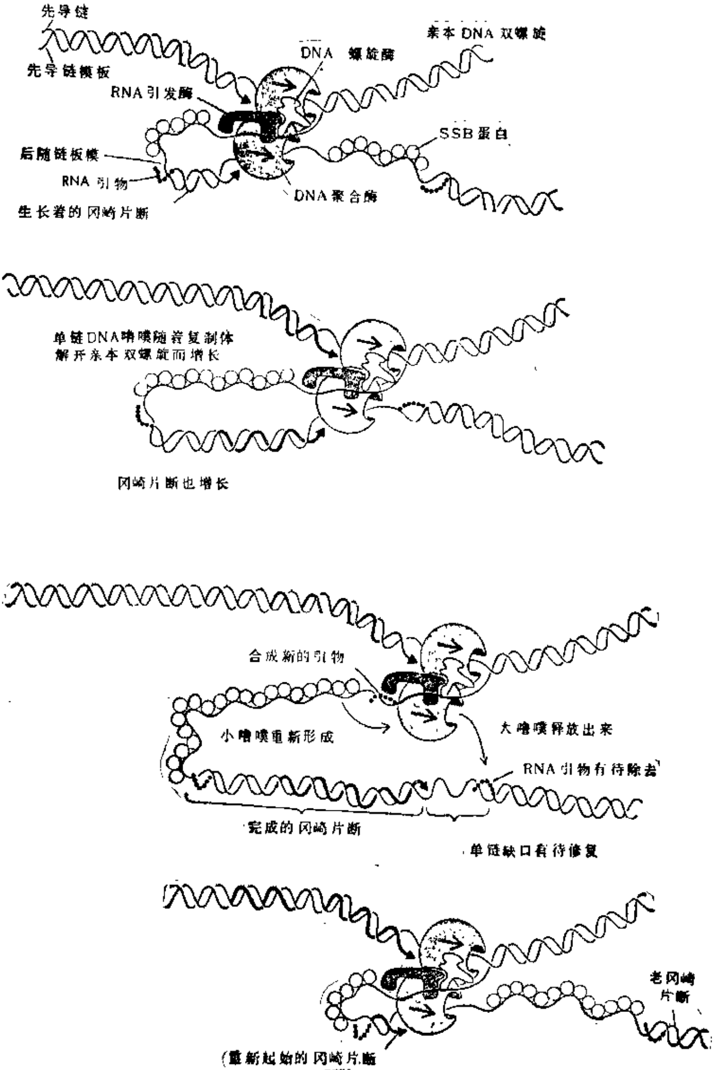



图 3-35 复制体与先导链-后随链同时复制模型

### 四、由复制体进行先导链和后随链的同时复制

在大肠杆菌染色体DNA进行复制时,DNA聚合酶全酶并不是单独起作用,而是与引发体,螺旋酶等构成与核糖体差不多大小的蛋白质复合体。这一复合体就叫做复制体 (replisome)。由于复制体的存在,先导链和后随链得以同时复制。现在虽然还未分离到完整的复制体,但是许多迹象表明,这一复合体确实是存在的。这些迹象主要有:

(1) 1988年 Kornberg 等人分离出 800 KDa 的 PolⅢ* 和 900 KDa 的 PolⅢ 全酶,其中各亚基均有两个分子,并证明是具有双活性位的非对称结构,这表明同一个 PolⅢ 全酶可能同时负责先导链和后随链的复制。

(2) 虽然先导链的合成是连续的,后随链的合成是不连续的,但是先导链上的 DNA 聚合酶 Ⅲ 全酶和后随链模板上的引发体均具有很高的进行性 (processivity),即不与 DNA 链中途解离,保证了先导链和后随链复制的迅速进行。事实上,先导链的合成总是领先一段,但也从来没有发现先导链跑得太远,而总是与后随链保持相对稳定的一段距离。我们已经知道,DNA 聚合酶 Ⅲ 全酶中的 $\beta$-亚基是负责进行性的亚基;而在引发体中并没有具有此功能的亚基。这就说明,引发体在 DNA 模板上的进行性是由于它与 DNA 聚合酶Ⅲ全酶的结合。

(3) 在 $\phi\times174$ 的复制中观察到引发体总是伴随着 DNA 嗜喋 (loop)。这些都表明,DNA 两条链的合成和引发有可能是由一个极为复杂的复制体进行的。图 3-35 就是由复制体对先导链和后随链进行同时复制的模型。

由图 3-35 可以看到,由于后随链的模板在 DNA 聚合酶Ⅲ全酶上绕转了 180° 而形成一个嗜喋,因此冈崎片断的合成方向能够与先导链的合成方向以及复制体的移动方向保持一致。随着冈崎片断的延长,这一嗜喋不断扩大。当冈崎片断合成到前一个片断的 5′ 端时,这一大嗜喋就释放出来。在此之前,先导链的合成又将另一部分后随链的模板置换出来,并在适于引发的位点上由引发体合成了新的引物,然后再形成一个小的嗜喋,进行新的冈崎片断的合成。需要指出的是,在这一模型中,引物的合成方向与引发体以及整个复制体移动方向相反,此处引发酶合成引物的细节还不清楚;然而与此模型相一致的是,后随链的引物均很短(见表 3-3)。由此模型不难看出,由于后随链的合成需要周期性的引发,因此其合成进度总是与先导链相差一个冈崎片断的长度。在复制叉前进的过程中,必须有 DNA 螺旋酶解开两条缠绕的模板链。起着螺旋酶作用的可能就是引发体中的 DnaB,它同时具有 ATPase 和螺旋酶的活性。此外,另一个螺旋酶 Rep 蛋白也可能参与解旋过程。螺旋酶并不能解决复制叉前进所造成的超缠现象,这必须由拓扑异构酶来解决。在 *E.coli* 体系中,主要是由 DNA 旋转酶承担这一任务的。

在绝大多数的 DNA 复制中,链延伸的速度大体上是恒定的,因此复制的起始是受到严格控制的。

## 第七节 复制的终止[54~61]

### 一、环形 DNA 复制的终止

在单方向复制的环形分子中,复制终点也就是它的复制原点。而在双方向复制的环形分

子中，有的有固定的终点，而大多数没有固定的终点，只是两个生长点的简单碰撞，而不管在什么位置。如 *E.coli* 及其噬菌体 λ。

环形分子复制的最后一步，必然是环连分子 (catenane) 的分离，这是由拓扑异构酶催化的反应，这方面目前人们所知不多。在 *E.coli* 系统中，据信是由 DNA 旋转酶(Eco 拓扑异构酶 II)所催化。

## 二、线形 DNA 复制的终止

线形 DNA 的复制显然不存在生成环连分子的可能性，但它面临着更为严重的问题。现在来让我们看一下噬菌体 T7 的线形 DNA 复制。T7 的复制原点在离一个端点 14.9% 处，其复制过程见图 3-36。

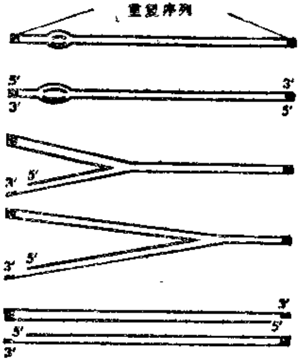


图 3-36　T7DNA 的复制过程中出现了隐缩的 5′ 末端

由图 3-36 可见，由于 RNA 引物的水解，两个子代分子中，各有一个 5′ 端缺少一段核苷酸，而没有一个目前已知的 DNA 聚合酶能填补。然而人们发现 T7 DNA 两端各有一段重复的数百个核苷酸，这叫做末端冗余。(terminal redundancy)。它意味着两个子代分子中的单链末端可以互补，多余的单链部分可以被 DNAase 除去。然后再由 DNA 连接酶连接起来。如图 3-37 所示。这样生成的二聚体叫做连环分子 (concatemer)。

现已知道，这种 DNAase 就是 T7 的基因 3 产物，分子量为 18 000。是个内切核酸酶，切断单链 DNA。基因 6 产物也可能参与连环分子的生成与处理。它是一个 5′→3′ 的外切核酸酶，其作用是水解细胞 DNA 以提供噬菌体 DNA 合成的核苷酸前体。由于 T7 DNA 聚合酶不具备 5′→3′ 外切酶活性，因而引物 RNA 的去除也是由基因 6 产物进行的。

生成的二聚体连环分子还可以通过复制产生四聚体、八聚体连环分子等。这一假说是 1972 年由 J.D.Watson 首先提出。在 T7 侵染细菌以后，分离到了巨大的 T7 连环分子，有力地支持这种假说。在任何时候，T7 DNA 分子都可以从这些连环分子上切下来(图 3-37)。

T7 的 DNA 复制反映了线形 DNA 在复制过程中所面临的问题。除了用连环分子的办法来解决之外，还有的线形 DNA 从一开始就采取环化的办法，例如 λ DNA 在进入寄主细胞以后就环化并成为超螺旋形式。有的噬菌体和病毒的线形 DNA (以及线形 RNA) 借助于蛋白质的介入来解决线形分子末端复制问题。这些噬菌体或病毒编码的蛋白质能与5′末端的碱基共价连结。腺病毒，噬菌体$\phi$29以及脊髓灰质炎病毒(一种 RNA 病毒)就采用这种手段。

腺病毒是哺乳动物的病毒，为线形双链DNA，长度大约为 36 000 bp。1986 年人们已弄清了人类腺病毒 Ad2 的 35 937bp 核苷酸序列。在腺病毒颗粒中 DNA 分子的每个5′末端都共价连接一个 55 KDa 分子量的蛋白质，称作末端蛋白质 (terminal protein)，简写作 TP。TP 以其丝氨酸羟基通过磷酸二酯键与 5′ 端的胞嘧啶共价连结。而正在复制的腺病毒新生 DNA 链的 5′ 端为 80 KDa 的蛋白质，这就表明 80 KDa 蛋白质参与了腺病毒 DNA 的复制。现在还不清楚在腺病毒的哪个成熟阶段，80KDa 蛋白质水解成 55KDa 蛋白质。现在通常把 80 KDa

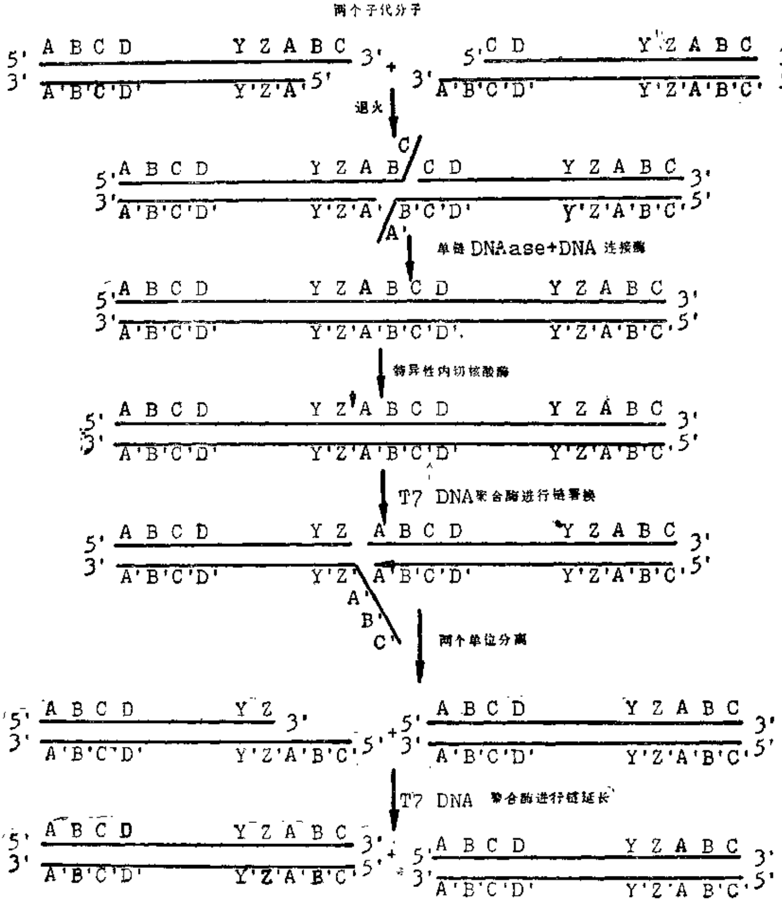



图 3-37 T7DNA 连环分子的生成以及完整单体分子的生成

蛋白质称作末端蛋白质前体，简写作 pTP。pTP 不但作为蛋白质引物解决了腺病毒 DNA 的复制起始问题，而且这一引发方式避免了线形 DNA 在复制结束后所面临的 5′ 末端隐缩的问题。

腺病毒两端都有反向末端重复序列，根据不同的血清型，其末端重复序列的长度介于 103 bp 至 162 bp。在两端各 50 bp 的范围内为其两个复制原点(图 3-38)。其中，第一个 C·G 碱基对和第 9～18 碱基对完全保守，核因子 NF1 的结合位点也高度保守。在头 25 个 bp 中，A/T 占 80%。

保守序列　　NF1 结合位点

(TP) -CATCATCAATAATATACCTTATTTTGGATTGAAGCCAATATGATAATGAG... 置换链
3'GTAGTAGTTATTATATGGAATAAAACCTAACTTCGGTTATACTATTACTC... 模板链

图 3-38　腺病毒的复制原点的序列

腺病毒在进行复制时，首先由核因子 NF1 与其结合位点相结合。然后，腺病毒 DNA 的末端在 72 KDa 的腺病毒单链 DNA 结合蛋白(DBP) 和 ATP 的存在时由 TP 催化产生末端解链。接着，140KDa的腺病毒DNA聚合酶(Adpol)能够催化pTP的丝氨酸残基与dCTP之间的反应，结果除去一分子焦磷酸并将 dCMP 以磷酸二酯键与pTP的丝氨酸残基共价相连；与此同时，C与模板 3′ 端的 G 以氢键相结合。这样就形成了起始复合物。末端解链的间接证据来自于体外试验，当 5′ 末端缺少TP时，必须先用 5′ 外切核酸酶切除置换链 5′末端的14个碱基，同样能够形成起始复合物。另一种可能的机制是先形成起始复合物，然后由于 pTP 与 TP 之间的某种相互作用而促进解链。解链必须有 DBP 和 ATP 存在，DBP 而且具有螺旋酶的性质。腺病毒的两个复制原点并不是同时起始复制，而是有先有后，采用链置换的方式进行。现在普遍认为，腺病毒的双链 DNA 中只有一个原点形成起始复合物而进行链置换；整个一条单链被置换出来以后形成“平锅柄”结构(panhandle)。然后再以同样的方式复制其互补链(图 3-39)。

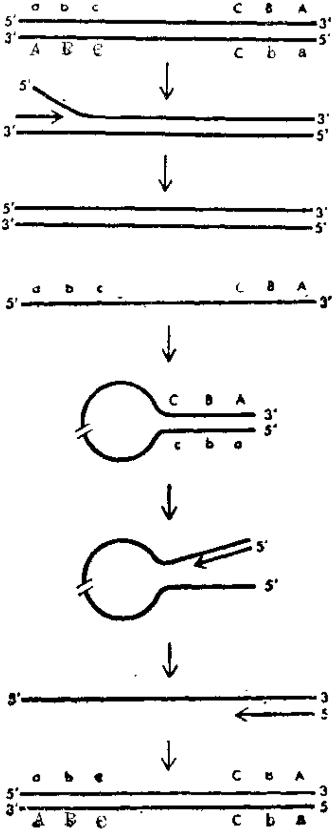


图 3-39　线形DNA(如腺病毒，$\phi29$)的复制过程。复制可以起始于任何一个原点，但是只有一个原点形成起始复合物并进行链置换反应。被置换出来的单链生成平锅柄结构，然后以同样方式复制其互补链。

腺病毒自身编码的三个蛋白质 pTP，Adpol 和 DBP 都位于同一个转录元中。腺病毒 DNA 的复制，除了需要寄主蛋白质 NF1 之外，还需要另一个寄主蛋白质——核因子 2(NF2)。NF2 是由 15KDa 亚基和 17KDa 亚基构成，具有拓扑异物酶 I 的性质。如果缺少 NF2，腺病毒 DNA 只能复制大约 20% 的长度便停顿下来。像腺病毒这样的线形 DNA 在复制过程中仍然需拓扑异物酶来解决超缠问题。可见，拓扑异构酶参与 DNA 复制，这是一个极为普遍的现象，或许是必不可少的。

枯草杆菌噬菌体 $\phi29$ 的情况与腺病毒十分相似。噬菌体自身编码 27 KDa 的末端蛋白质和 DNA 聚合酶是这一复制过程必不可少的。此外，

还需要其他一些由寄主编码的蛋白质。

灰质炎病毒，其5′末端也通过磷酸二酯键与VPg蛋白的酪氨酸残基相连。VPg只有22个氨基酸。但人们已经发现其前体，分子量为12 KDa，且与细胞膜相连。12KDa的前体可能就是参与引发过程的蛋白质。负责灰质炎病毒RNA复制的RNA聚合酶称为RNA复制酶，它像DNA聚合酶一样，也是严格依赖于引物的。

## 第八节　真核生物的DNA复制[57,62~78]

在上一节中，我们已经从解决线形DNA复制终止问题的角度讲解了一种真核生物病毒——腺病毒的DNA复制。本节我们将要讲解真核生物的复制原点，复制元和复制元族；真核生物的DNA聚合酶和引发酶；作为真核生物DNA复制模型的SV40的复制原点和大T抗原以及复制过程；真核生物染色体末端的DNA的复制；真核生物复制过程中的核小体结构。

### 一、真核生物的复制原点，复制元和复制元族

在*E.coli*中，DNA复制速度大约是$10^5$bp/分，而真核细胞的DNA聚合酶活性要低得多，复制速度大约为500～5000 bp/分。这样，如果按典型的哺乳动物染色体DNA（比*E.Coli* DNA大50倍）来计算，则真核DNA的复制时间就会是*E.Coli*的1 000倍即约一个月的时间。事实上，真核生物DNA复制时间一般为几个小时。这是通过从许多复制原点同时开始并双向复制而实现的。用放射自显影方法在哺乳动物染色体上看到许多复制泡，每个复制泡都有固定的起点（复制原点），然后双向伸展，与相邻的复制泡会合。这样一段段的DNA就称为复制单元，简称复制元（replicon）。复制元的大小是不均一的，从13 000 bp～900 000bp不等。不同生物的复制元大小当然不会相同；就是同一种生物在不同的生长条件下，复制元的大小也不同。生长快的时候，复制元就小。例如果蝇大约有5 000个复制原点，每个复制元平均大小约30 000bp；而在卵受精后，复制起点增加到大约50 000个，仅需3分钟就可以将整个基因组复制完毕。其调节机制目前毫无所知。几个邻近的复制元可组成复制元族，每个复制元族少至两个复制元，多至250个复制元。不同的复制元族在复制起始的时间上有先有后，而同一复制元族内各复制元基本上是同步的。真核生物的复制原点的DNA序列并无固定的模式，但大多包含一个富含AT的序列，可能还有一个特异性蛋白质的结合位点。

### 二、真核生物的DNA聚合酶和引发酶

与真核细胞中多起点复制相应的特点是细胞中DNA的聚合酶分子数目多。在*E.Coli*中，只有polⅢ 10～20个分子，而一个典型的动物细胞含有DNA聚合酶$\alpha$ 2 000～6 0000个分子。DNA聚合酶$\alpha$可能是真核生物的主要的DNA聚合酶，此外还有DNA聚合酶$\beta$，$\gamma$和$\delta$。现将四种DNA聚合酶的性质总结于表3-6。大多数真核生物DNA聚合酶都只有聚合活性，而不像细菌DNA聚合酶那样除了聚合活性外还有外切酶活性。而新近发现的DNA聚合酶$\delta$不但有聚合酶活性，而且还有3′→5′外切酶活性。DNA聚合酶$\alpha$与DNA引发酶结合得相当紧密，甚至在用DNA聚合酶$\alpha$的抗体进行亲和层析时，这个紧密的复合物也一同被吸附；但DNA引发酶可以被50%乙二醇（ethylene glycol）所洗脱。因此前几年人们认为DNA聚合酶$\alpha$兼具引发酶的活性是不恰当的。实验表明，DNA聚合酶$\alpha$和DNA引发酶的复合物

是复制体系所必需的。DNA 聚合酶 α 和 δ 在体外都可以被一种霉菌的抗菌素 aphidicolin 所抑制，在体内该抗菌素则抑制细胞 DNA 的复制。实验表明，在细胞 DNA 合成期(S 期)，DNA 聚合酶 α 的含量急剧增加。而对于 DNA 聚合酶 δ，控制细胞周期的蛋白质 cyclin 则是它不可缺少的辅助成分(亚基?)。据此，人们推测，真核生物的 DNA 复制是由 DNA 聚合酶 α 和 δ 协同进行的。DNA 聚合酶 β 在细胞周期中含量变化不大，可能与损伤 DNA 的修复有关。DNA 聚合酶 γ 的功能主要是负责线粒体 DNA 的复制，而它在核内的功能（如果有的话）还不清楚。

表 3-6 哺乳动物的四种 DNA 聚合酶

| | DNA聚合酶α | DNA聚合酶β | DNA聚合酶γ | DNA聚合酶δ |
|---|---|---|---|---|
| 位　置 | 核　内 | 核　内 | 线粒体及核内 | 核　内 |
| 活性比例 | ~80% | 10~15% | 2~15% | 未计算 |
| 亚基数 | 多亚基 | 单亚基 | 单亚基 | 多亚基 |
| 分子量 (KDa) | 110~120 | 45 | 60 | ? |
| aphidicolin 抑制反应 | + | − | − | + |
| BUPdGTP 抑制反应 | + | − | − | − |
| DNA 引发酶活性 | + 因与引发酶紧密结合 | − | − | − |
| 3′→5′ 外切酶活性 | − | − | − | + |

前面，我们介绍了真核生物的DNA聚合酶。现在，我们还必须了解真核生物 DNA 引发酶的性质。引发酶分子量大约 50~60 KDa。首先，正如上面所说，引发酶能够与 DNA 聚合酶 α 形成紧密的复合物。这种复合物不但是复制所必需的，而且与结合于复制原点的特异性蛋白质构成了复制体系的种族特异性。例如，来自人类 HeLa 细胞的复制体系，如果用 DNA 聚合酶 α 抗体去除体系中的 DNA 聚合酶 α 和引发酶的复合物，则复制体系的其余组分不能支持 SV40 复制原点在体外的复制(在大 T 抗原存在时)。如果加入人类 HeLa 细胞中的这种复合物或猴子 COS 细胞中的复合物，均能恢复复制能力。而加入小鼠或牛的复合物以及 *E.coli* DNA 聚合酶Ⅰ或聚合酶Ⅲ全酶均不能奏效。这可能是该复合物与大T抗原之间的蛋白质-蛋白质的特异性作用有关(见下一小节)。其次，真核生物的 DNA 引发酶的引物合成方式较为特别。它的作用是阵发式的，每次作用能拷贝 6~15 个核苷酸。当 DNA 聚合酶 α 缺少时，首次阵发性合成产物可以被下几次阵发性合成所延长成为很长的多聚核苷酸。但是在有活性的 DNA 聚合酶 α 和足够的 dNTP存在时，引物仅是第一次阵发性合成产物。最后，DNA 引发酶不但能利用 rNTP 为合成前体，而且还利用 dNTP 为合成前体。在一次阵发性合成的引物中，第一个核苷酸总是 rA；而以后的核苷酸要么都是核糖核苷酸，要么都是脱氧核糖核苷酸，而不能是二者都有。由于引发酶能够合成少量的脱氧核糖核苷酸，因此精确确定 DNA 聚合酶 α 何时接替引发酶是很困难的。然而，引物一般只有 6~15 核苷酸，对于 SV40 体外 DNA 复制情

况的分析，发现它的引物仍然主要为6～9核苷酸的RNA。

### 三、SV40的大T抗原与复制原点

SV40仍然是人们目前研究真核生物DNA复制的主要模型。它的双链环形DNA有5243 bp，在寄主细胞内能够形成具有核小体结构的微染色体。SV40所编码的蛋白质中，只有一个大T抗原参与其DNA复制，其余的复制机构均由寄主的蛋白质构成。因此对于SV40复制机制的探讨是很具有吸引力的。大T抗原四聚体在复制原点区有三个结合位点（见图3-40），同时还具有ATPase活性和螺旋酶活性。此外，最近还发现，大T抗原还具有某种特异的蛋白质-蛋白质相互作用，因为有一种突变的大T抗原保留了结合复制原点的能力和ATPase活性，而不能促进SV40 DNA的复制；还有一种突变的大T抗原失去了结合复制原点的能力，但能在体外复制系统中能够与野生型大T抗原相竞争而影响到SV40的复制。因此大T抗原的作用可能与*E.coli oriC*复制体系中DnaA的作用相似。对于寄主所提供的复制因子，除DNA聚合酶$\alpha$和DNA引发酶之外，其余的众多因子还没有很好地进行分离和研究。目前已有证据表明，160 KDa的寄主核因子NF1也参与了SV40的DNA复制。

SV40的DNA复制原点，最低限度得保留5 208到29这样一段65 bp的DNA序列，才不至于完全丧失DNA复制能力。但在这种情况下复制能力只相当于野生型的40～50%。如果要保留≈100%的复制能力，则必须从5 192到30这样一段82 bp的DNA序列。其中包括了大T抗原结合位点Ⅰ和Ⅱ，连续的16 bp A/T序列（野生型为17 bp）以及左向复制的先导链（合成方向与早期基因转录的方向相同）和后随链的最迟RNA→DNA转变位点（图3-40）。82 bp的*ori*序列是根据15种不同的缺失突变*ori*序列的体外复制情况而确定的。结合位点Ⅰ是大T抗原的最主要的结合位点，但是它可以容忍相当大的部分缺失而对复制功能影响较小；而结合位点Ⅱ以及其余的重要序列的很小缺失都会完全丧失复制功能。而SV40 *ori*序列的复制是绝对依赖于大T抗原的。这里的脱节情况目前尚无一种有充足理由的解释。

### 四、SV40以及其他真核生物的DNA复制过程

根据核苷酸水平上RNA引物与DNA的转变位点的研究（图3-41）人们发现SV40左向复制的先导链（方向与早期基因转录方向相同）的DNA5′端有10个主要位点，其中最主要的两个位点（5 239，12）均位于27 bp的迴文序列内。在27 bp的迴文序列内有大T抗原结合位点Ⅱ，早期m RNA的起始转录点。这一区域和富含A/T区域对于SV40复制均是不可缺少的。此外，在6个$G_3CG_2Pu_2$序列处也是常见的引发位点。已知转录因子SP1结合在这些序列上，其促进复制的机制还不清楚。而在右向复制的链（方向与晚期基因转录方向相同）上，DNA 5′端有6个主要位点。值得注意的是右向复制的6个主要转变位点与左向复制的10个主要转变位点完全不重叠，双方只有很次要的转变位点在5 210～5 211这样极其狭窄的范围内重叠（图3-41）。而且这些右向复制的转变位点大部分位于必需的原点序列之外，其缺失并不影响复制的进行。

由上可见，SV40 DNA从复制原点开始的两个方向的复制并不是同时开始的。可能由于大T抗原结合于原点序列，以其螺旋酶活性促进负超螺旋状态下的17bp的A/T序列的解链，同时由单链DNA结合蛋白结合于解链部分。这一解链区域可能以动态方式向两侧传播，暴露出适于引发的单链DNA序列。然后由DNA引发酶合成引物，特别是优先在27 bp的迴文

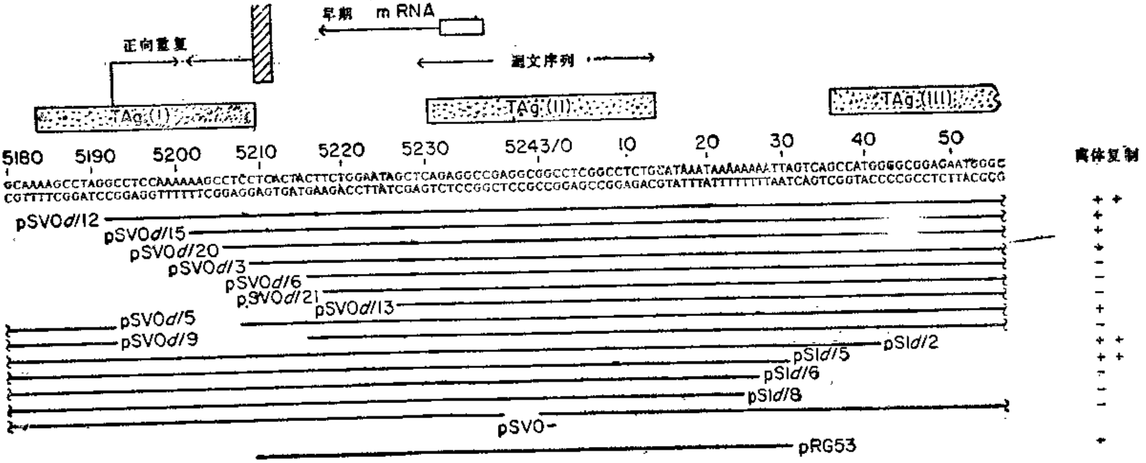



图 3-40　SV40 DNA 复制原点及其两翼的部分序列。在序列上方，TAg(Ⅰ),TAg(Ⅱ),TAg(Ⅲ)分别表示大T抗原的三个结合位点；←□ 表示早期基因转录起始点；////// 表示位于 5 210-5 211 的转变位点(见正文所述)。序列下方的15条横线表示在质粒 pSVO,pS 1,pRG53 中复制原点的序列缺失情况；右侧的正负号表示这些质粒在体外复制体系中的复制能力。(＋＋)表示具有 100%的野生型复制能力，(＋)表示有40～50%的复制能力，(－)表示缺乏可检测的复制能力

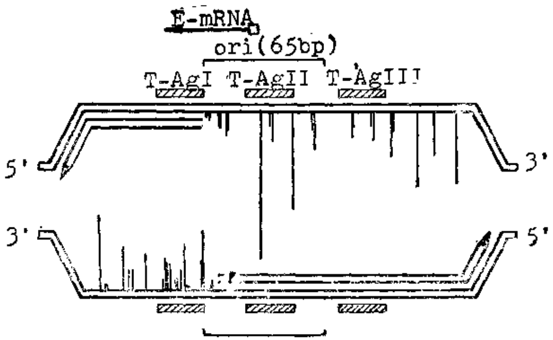



图 3-41 SV40的最低限度复制原点以及左、右向复制中RNA→DNA 的转变位点的不重叠性。*ori*(65bp)为最低限度复制原点，T-Ag(I)，(Ⅱ)，(Ⅲ)分别表示大T抗原的结合位点。E-mRNA 表示早期基因的转录产物，箭头表示其转录方向。每个模板上的垂直线表示引物 RNA→DNA 的转变位点，垂线的长短代表实际测量出来各转变点出现几率的大小。

序列内进行引发，从而开始左向复制的先导链的合成。当先导链合成至原点序列的左侧，才暴露出适合右向合成的模板序列，由另一个复制体引发并合成右向复制的先导链。在细胞内也有可能是先作为左向复制的后随链首先被引发与合成，然后由另一个复制体进行延伸，并开始其后随链的引发与合成。这个过程示于图 3-42。这样，SV40 DNA复制的起始就像 *E.coli* 一样，两个方向的复制叉对于原点序列来说是不对称的，作为双向复制的对称点出现在复制原点序列的左半边。

根据引发位点和引物序列的分析，在 SV40 DNA 复制过程中，先导链和后随链的引发都是由 DNA 引发酶进行的，引物 RNA 的长度均为 6～9 核苷酸。aphidicolin 是 DNA 聚合酶 $\alpha$ 强有力的抑制剂，而对 DNA 引发酶不起作用。当复制体系中有 aphidicolin 存在时，能够抑制复制叉的延伸，而不能抑制复制原点处有限的 DNA 合成。这就足以说明，复制原点处的引物(即先导链的引物)是由 DNA 引发酶合成的。

SV40 DNA 复制过程中，部分复制的中间产物呈蝴蝶形(图 3-43)。在蝴蝶形分子中，正在复制的部分成为唵噗，而尚未复制的部分仍然为超螺旋。SV40 像噬菌体 λ 一样，也没有固定的复制终点，只是两个复制叉的简单碰撞。复制完毕或接近完毕时，必然形成环连体分子。在拓扑异构酶Ⅱ的作用之下发生去环连作用，这一过程可能主要发生在复制结束之前。

SV40 的 DNA 复制有可能代表着真核生物 DNA 复制的主要过程。首先由一个特异的蛋白质识别复制原点并与之结合。然后由螺旋酶促进负超螺旋状态下的复制原点的解链，而单链 DNA 结合蛋白维持了解链状态并使解链区域有可能以动态方式向左右转移。由 DNA 聚合酶 $\alpha$ 和 DNA 引发酶与原点序列上的特异蛋白质相互作用，从而引发了一个复制叉的先导链的合成；然后再引发另一个先导链的合成(并不需要在原点序列内，也不需识别原点的特异

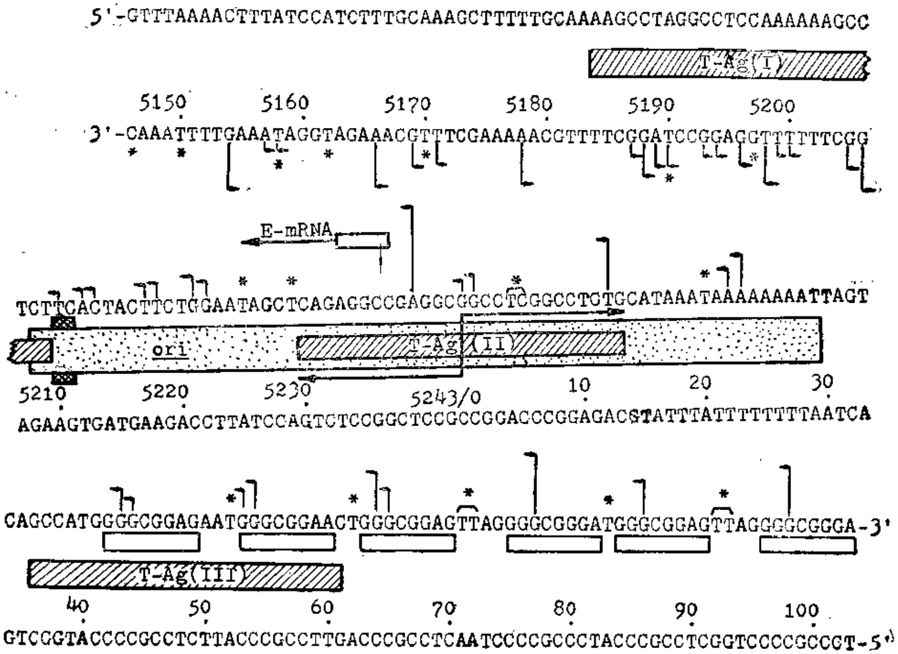
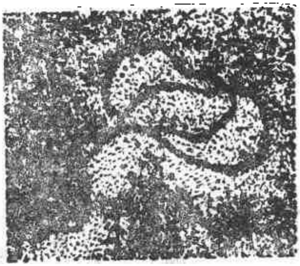
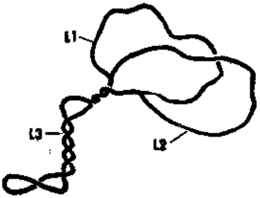



图 3-42　SV40 DNA 复制起始的可能方式　本图所示 *ori* 为 SV40 DNA 复制所必需的最低限度的序列。大 T 抗原的三个结合位点用斜线长方块表示。SP1 的六个结合位点用空白长方块表示。星号表示引物 RNA 合成的起始位点。带箭头的竖线表示共价连接于引物 RNA 的新生 DNA 的 5′ 端位点，箭头表示新生 DNA 链的延伸方向，竖线长短表示在此处起始 DNA 合成的相对频度的大小。网状小方格表示两个方向的 DNA 合成的最后转变位点 (5 210–5 211)，即左向 DNA 合成均起始于此位点的右面，右向 DNA 合成均起始于此位点的左面。


(a)



(b)

图 3-48　SV40 复制过程中的蝴蝶形中间体

(a) 复制早期的电子显微镜照片　(b) 电子显微镜照片图解，L1，L2 为已经复制的部分，L3 为尚未复制的部分，且呈超螺旋构象。

的蛋白质)。后一个先导链的引物或许是 RNA，或许是第一个复制叉的第一个冈崎片断。随着复制叉的前进，还需要拓扑异构酶解除复制叉前进所造成的超缠现象。在酵母中起这种作用的可以是拓扑异构酶 I 或拓扑异构酶 Ⅱ；而在 SV40 中为拓扑异构酶 Ⅱ。真核生物染色体 DNA 从许多复制原点开始复制。当相邻的两个复制元的相向的两个复制叉随机会合时就完成了这一段 DNA 的复制，可能并不存在固定的终点序列。然而，真核生物染色体末端 DNA 的复制却不那么简单。

### 五、真核生物染色体末端 DNA 的复制

真核生物同样面临着线形 DNA 复制结束时所产生的 5′ 末端隐缩的问题。这一问题的解决有赖于端粒的结构和能够识别或结合端粒的蛋白质和酶。关于端粒蛋白质的研究正在进展之中，已发现的几种端粒蛋白质已在第二章中讲过。Blackburn等人首先发现四膜虫端粒中存在一种端粒酶 (telomerase)，这种酶能够将四膜虫端粒结构中单链尾巴 5′TTGGGG3′ 延长，延长的部分仍然是 5′TTGGGG3′。事实上，各种端粒中这一富含 G 的序列总是突出 12-16 个核苷酸。后来又发现原生动物游仆虫 ( euplotes) 和尖毛虫的端粒酶也能延长其富含G序列 5′TTTTGGGG3′。这种酶是一种核糖核蛋白，由 RNA 和蛋白质构成。1990 年，Blackburn 等人又证明了端粒酶中的RNA是富含 G 序列的模板。例如游仆虫的端粒酶 RNA 含有 5′CAAAACCCCAAAACC3′ 这样的序列，对于拷贝出富含 G 序列所必需的部分为 5′CAA-AACCCCAAAA3′。这样，端粒酶实际上是一种逆转录酶，与还原病毒逆转录酶的差异只在于这个模板是在酶分子内。端粒酶中的 RNA 可能还有别的作用。这一富含 G 的单链尾巴的长度是如何控制的以及这一单链尾巴在复制中的功能都还不清楚。这种单链尾巴可能弯转过来作为引物复制 5′ 末端。另一种可能性是某种端粒结合蛋白有可能像腺病毒 pTP 那样结合于最末端的单链重复序列，并作为蛋白质引物复制 DNA 的 5′末端。

### 六、真核生物复制过程中的核小体结构

真核生物细胞核的体积是很小的，而复制过程完全局限于核内进行，因此 DNA 必然是处于高度压缩状态下进行复制的。而复制时必然涉及到核小体的转移与分配，因为 DNA 是半

保留复制。由于空间狭小，复制可能是分区进行的，这就是各复制元族不同步的原因。现在人们已经用溴尿嘧啶脱氧核苷酸代替胸腺嘧啶核苷酸渗入到DNA，再用溴尿嘧啶脱氧核苷酸的抗体进行免疫杂交而证实了染色体的分区复制。

第二章中我们已经讲过，组蛋白八聚体是以全保守的方式传给子代分子的，现将得到这一结论的实验过程叙述如下。

组蛋白的合成是在细胞核中与DNA复制同步进行的，因而细胞并不含有明显多余的组蛋白分子。其耦联的机制目前还不清楚。这个事实对于我们证明组蛋白八聚体的解离与否却有很大的帮助。老的八聚体是否解离以及如果解离又如何与新合成的组蛋白重组？这里主要有三种可能性(图 3-44)。

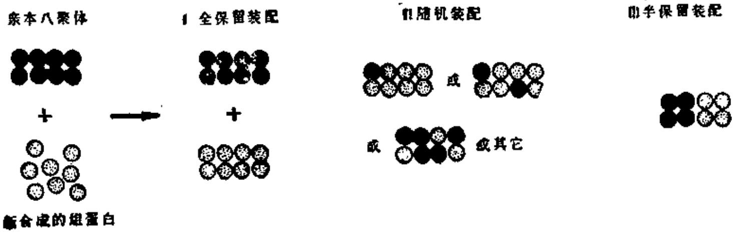



图 3-44　子代八聚体的可能的组合方式

将细胞置于轻培养基 ($^{12}C$, $^{14}N$, $^{1}H$) 中生长很长时间。然后将细胞转移到重培养基 ($^{14}C$, $^{15}N$, $^{3}H$) 中生长约1小时，分离组蛋白八聚体，在甲酸饱和的密度梯度中进行平衡离心。如果是全保留装配的话，则得到一个高密度的放射性带；如果是随机的话，则得到一个高密度到低密度的弥散的放射性带；如果是半保留的话，则在中等密度处得一个单一的放射性带；结果表明，组蛋白八聚体的装配是全保守的。

这些老的八聚体和新的八聚体在两条子代DNA分子上是如何排列的呢？同样有三类可能性(图 3-45)。

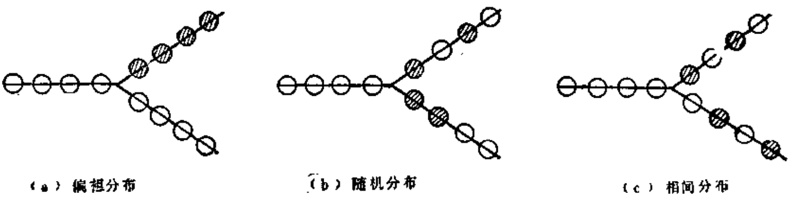



图 3-45　新、老八聚体的可能排列方式

用上述密度梯度平衡离心的办法分析交联的相邻八聚体 (16 单体聚合物)同样可以证明第 I 种分布是正确的。

用蛋白质合成的抑制剂放线菌酮 (cycloheximide)可以得到同样的结论。将这种抑制剂加到正在生长的细胞中，在此之前开始的一轮 DNA 复制将继续下去。上面我们已经谈到过组蛋白是现用现造，在加入蛋白质合成的抑制剂后，就没有新生的组蛋白八聚体了。这样在两个

复制叉的两边，有一个分枝的 DNA 是裸露的。另一个分枝继承了亲代的组蛋白八聚体。用电子显微镜可以看到这种现象(图 3-46)。这一结果表明，组蛋白八聚体在 DNA 复制时并不离开亲本 DNA 链。

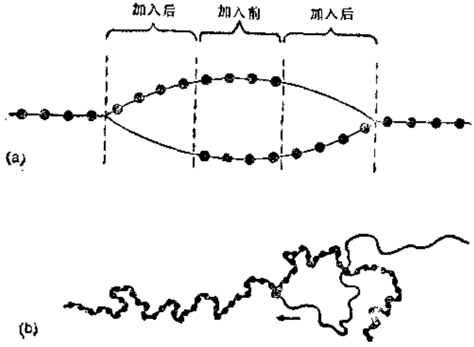



图 3-46 放射菌酮抑制实验

(a) 模式图 (b) 电子显微镜照片及解释图

最后的问题是，亲代的组蛋白八聚体到底与哪条链相结合？迄今，人们只知道一种动物病毒 SV40，它的老八聚体是与先导链相结合的(SV40 也是有核小体结构的)。如果一个复制单元从原点开始双向等速复制，则在子代分子的这一段序列上的组蛋白八聚体半截是老的，半截是新的(图 3-47)。

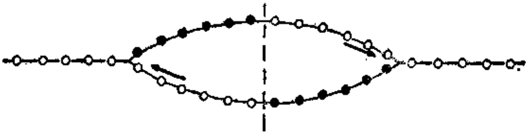



图 3-47 新(●)、老(○)八聚体在子代分子上的分布。箭头表示先导链前进方向

## 第九节 复制的调控[79-83]

许多现象表明，DNA 的复制是受到调控的。例如在葡萄糖作为唯一碳源辅以一些无机化合物时，*E.coli* 细胞在37℃时分裂一次需要 45 分钟，在琥珀酸为碳源时需 70 分钟，在很差的碳源中可以长达 10 小时。然而对于 DNA 复制来说，都为 40 分钟左右，即使细胞分裂需要 10 小时的情况下，也仅仅增加几分钟。更为奇怪的事情是，在非常丰富的培养基中，结胞分裂时间只有21分钟，然而 DNA 复制仍然需要 40 分钟。这可以用成熟前起始(premature initiation)加以解释(图 3-48)。这种现象一般可以用电子显微镜观察到。

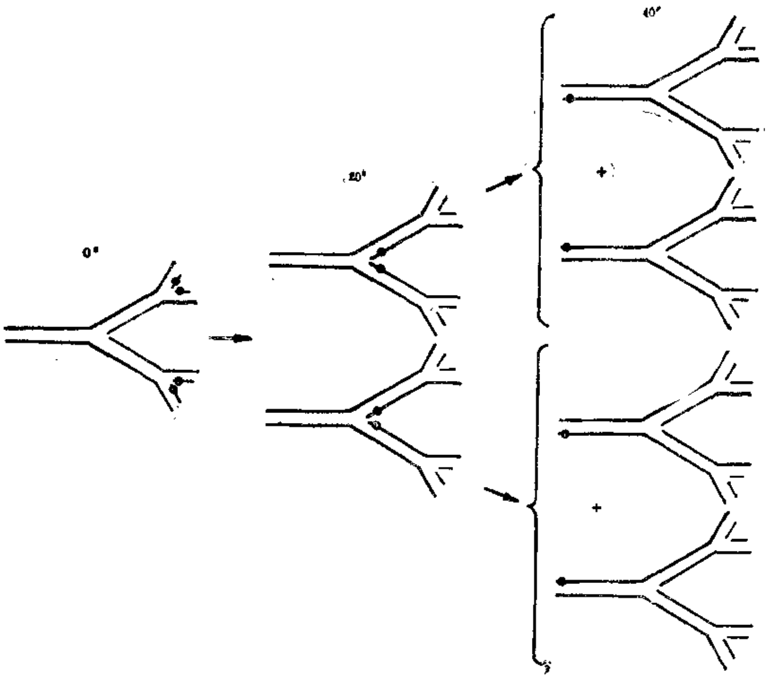



图 3-48　成熟前起始图解

图中小圆点作为标题以示0'、20'、40'各个时刻复制的进展情况，由图可见，分子加倍只需20分钟，但一个分子复制完毕仍需要40分钟。

然而，关于 DNA 复制的调控，人们仍然很少了解，可能涉及到许许多多的专一性蛋白质因子(如 *E.coli* 的 DnaA，SV40 的大T抗原等) 和至少一类RNA分子。细胞中蛋白质和 DNA 总量的比例可能也是一个很重要的因素，因为氨基酸饥饿不但造成蛋白质合成的抑制，而且也造成 DNA 复制起始的抑制，然而又不知为什么能够允许正在复制中的 DNA 的复制完毕。

从复制原点起始一轮 DNA 复制，特异性蛋白质因子对原点的识别和结合是关键的一步。当 DNA 复制起始以后，这种可为复制机构所利用的状态发生了改变，或者是 DNA 被修饰(如甲基化和去甲基化等)，或者是特异性蛋白质因子被修饰(磷酸化和去磷酸化等)。这样，使得在一个细胞周期中，复制原点只能被使用一次，而不是多于一次。前面我们已经讲过，*E.coli* 复制原点中 GATC 序列的保守性和 *dam* 基因对于复制起始的重要性。近来人们发现，SV40 复制原点上结合的大 T 抗原四聚体起初是没有磷酸化的；复制一旦起始以后，大 T 抗原就被磷酸化从而离开了复制原点。这也同样保证了复制原点的一次性使用。RNA 也参加了复制起始的调控。人们比较清楚的就是大肠杆菌质粒 ColE1 的复制。ColE1 的复制需要一个 RNA 引物。RNA 聚合酶从复制原点上游方向 555 bp 处起始转录，这个转录物就称为 RNA 2。当

这个 RNA 2 延伸至复制原点区域时，由RNAaseH将 RNA 与 DNA 模板形成的杂合双链中的 RNA 切断，从而产生 3′-OH 末端。然后，DNA 聚合酶以此为引物合成 DNA(图 3-49)。

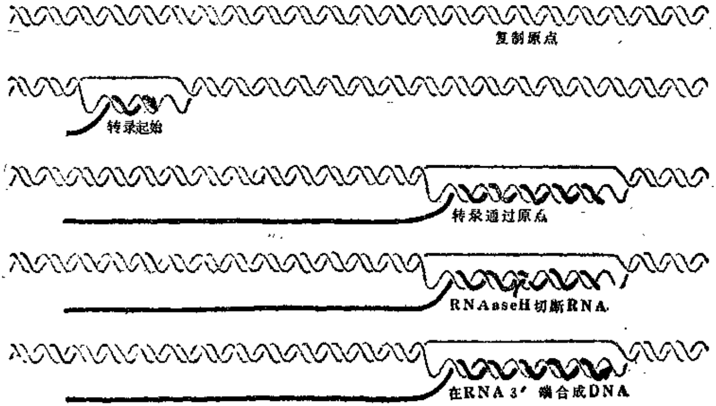



图 3-49 通过 RNAaseH 切断 RNA 2 而产生具有 3′-OH 的引物，从而起始 ColE1 的复制

在 ColE1 复制原点上游方向 445 bp 处(相当于 RNA2 的第111个碱基的位置)，起始转录另一个 RNA 分子，其转录方向与 RNA2 相反。这个 RNA 分子称为 RNA1。这个反义的 RNA1有 111 个碱基与 RNA2 的 5′ 末端互补，从而改变了 RNA2 的二级结构，使之不能为 RNAaseH 所识别（人们通常认为 RNAaseH 只识别 DNA-RNA 杂合双链，而不识别 RNA 的二级结构。因此其细微机制仍有待于作进一步的研究)。于是，RNA2 的转录继续进行，不能在复制原点区域产生引物。ColE1 复制原点下游方向不远处有一个编码 63 个氨基酸的蛋白质基因。这个蛋白质称为 Rom 蛋白(有的文献上称作 Rop 蛋白)。有的研究者发现，Rom 蛋白在 RNA1 存在时能够促进 RNA2 在合成到 100～220b 以后终止，从而不能达到复制原点处，也就不能产生复制所需的引物。最近有人证明，当RNA2长度介于200～360b时，Rom蛋白有促进 RNA1 与 RNA2 结合的能力。当 RNA2 长于 360b 时，RNA1 虽然也能够与 RNA2 结合，但并不影响引物的形成。体外实验表明，RNA2 的二级结构将影响它与 RNA1 的结合；而 RNA2 的长度不同，其二级结构也不同。因此，体内 RNA1 与 RNA 2 的有效结合是发生在 RNA2 转录过程的特定时刻。

复制起始的控制与质粒(或染色体)的拷贝数的控制是完全一致的。目前关于拷贝数的控制机制虽然有若干种模型(阻遏物控制，反义 RNA 控制等)，但实际情况可能远为复杂。每个细胞中只有 1～2 个拷贝的质粒称为严谨型质粒(stringent plasmids)或低拷贝数质粒；每个细胞中有 10～100 个拷贝的质粒称为松驰型质粒(relaxed plasmids)或高拷贝数质粒。质粒的自私特性即不相容性 (incompatibility) 也是与复制起始的控制密切相关的。具有相同或者相似阻遏物的两种质粒常常不能稳定地存在于同一个细胞中，经过若干代的培养，只含有同一种质粒的细胞越来越多，而含有两种质粒的细胞相对减少(图 3-50)。由此可见，质粒的不相

容性是两种质粒具有相同或者相似的阻遏物以及两种质粒复制的随机选择的结果。有些质粒能够稳定地存在于同一细胞中，它是因为它们控制拷贝数的机制完全不同。例如F质粒和ColE1质粒二者是相容的，也就是说，它们分别属于不同的不相容性组(different incompatibility groups)。

现在已有证据表明，复制起始的调控不但与拷贝数的控制有关，而且还与这些拷贝在子代细胞中的均等分配有关。

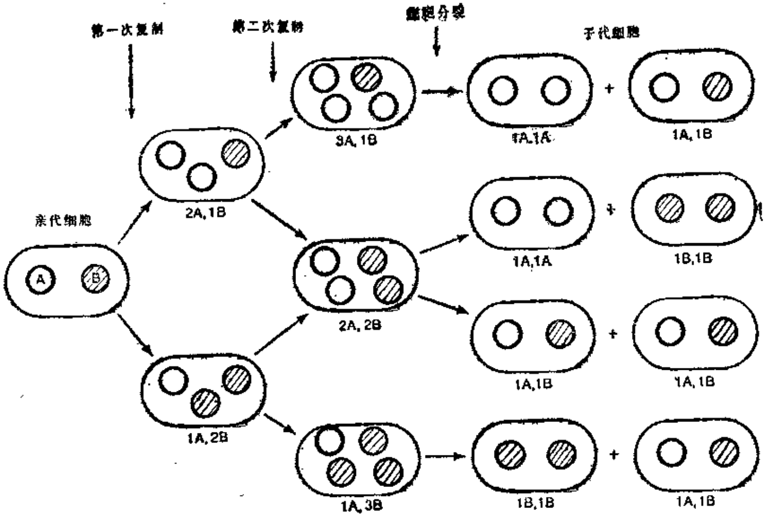



图3-50 相同或相似的阻遏物以及复制的随机选择造成了两种质粒的不相容性

## 提 要

Watson和Crick的DNA右手双螺旋理论本身就暗示着双链DNA复制的基本模式，也就是Meselson和Stahl同位素标记实验所证实了的双链DNA的半保留复制。然而具体的DNA复制机制却要复杂得多。

DNA复制通常是从特定的复制原点开始，大多数双向进行，也有少数单向的或以不对称的双向方式进行的。根据先导链合成的起始，可以将DNA合成分为从新起始(复制叉式复制)和共价延伸(滚环复制)两大类型。

大肠杆菌有三种DNA聚合酶，其中DNA聚合酶Ⅲ是复制过程中的主要聚合酶。该酶是由10种22个亚基构成的分子量为900KDa的非对称聚集体，这就暗示着这一聚集体可能同时进行先导链和后随链的复制。DNA聚合酶Ⅰ在DNA复制过程中起重要作用。DNA聚合酶Ⅰ也是实验室中最常使用的酶之一。DNA聚合酶以四种脱氧核苷三磷酸作为前体。连接酶和拓扑异构酶也是DNA复制过程中所不可缺少的。

在双链DNA的复制过程中，先导链的合成是连续的；而后随链的合成是不连续的，先合

成小的片断(冈崎片断),而后连接起来。每个冈崎片断的合成都需要一个 RNA 引物,这种 RNA 引物在大多数情况下都是由引发酶合成的。有时是由 RNA 聚合酶合成的。在合成引物之前,常常需要 RNA 聚合酶在复制原点附近进行转录(转录激活)。周期性起始的冈崎片断需要 DNA 聚合酶Ⅰ和 RNAaseH 去除引物,DNA 聚合酶Ⅰ进行填补,最后由 DNA 连接酶连接到已经合成的后随链上。

复制原点是 DNA 复制的起始场所,而且需要有特异的蛋白质因子与之结合。大肠杆菌的*oriC* 至少包括 245 bp;其他革兰氏阴性菌的复制原点亦很相似。不同的复制原点需要不同的蛋白质因子,即使同一细胞体系中的不同复制元也是这样,例如,*oriC* 需要 DnaA,而 $\phi\times$174 的复制就不需要DnaA。大肠杆菌的DNA聚合酶Ⅲ全酶是具有双活性位的非对称结构,这暗示着它可能同时进行先导链和后随链的合成。这就需要DNA聚合酶Ⅲ全酶与引发酶为中心的引发体构成更大的复制体。这样,从原点开始的一个复制叉的第一个冈崎片断,并不是相反方向复制叉的事实上的先导链;相反方向复制叉的启动需要装配另一个复制体。

环形双链 DNA 一般没有固定的复制终点。复制将近结束时,由拓扑异构酶Ⅱ进行解环连。线形双链 DNA 的末端复制面临着很大的问题,有的通过末端冗余和形成连环体来解决,如T7 DNA;有的利用蛋白质作为复制起始的引物,如真核生物的腺病毒。

真核生物有四种 DNA 聚合酶,其中 DNA 聚合酶 $\alpha$ 可能是主要的聚合酶。真核生物 DNA 聚合酶活性很低,而以众多的复制原点作为补偿。真核生物具有性质独特的 DNA 引发酶。真核生物的复制原点也富含 A/T 碱基对,同样需要特异性的蛋白质结合因子。SV40 的复制原点至少包括 82 bp,结合因子为大 T 抗原。真核生物可能具有与大肠杆菌类似的复制起始机制。真核生物的线形 DNA 末端可能由特异的端粒蛋白质负责末端 DNA 的复制。在 DNA 复制过程中,原有核小体蛋白质是全保守地结合于先导链的模板。

DNA 复制与细胞周期紧密相关,因而它必须处于严格的调控之下。人们在这方面所知甚少,可能涉及到许许多多的蛋白质因子和至少一类 RNA 分子。

## 参 考 文 献


1. Watson J. D. and F. H. C. Crick (1953) "Genetic Implications of the Structure of Deoxyribonucleic Acid." *Nature*, 171: 964～967.
2. Meselson, M. and F. W. Stahl (1957) "The Replication of DNA in *E. coli*." *Proc. Nat. Acad. Sci.*, 44: 671～687.
3. Kornberg, A. (1979) "Aspects of DNA Replication." *Cold Spring Harbor Symp. Quant. Biol*, 43, 1～9.
4. Cairns, J. (1966) "The Bacterial Chromosome." *Sci. Amer.*, 214: 36～44.
5. Gilbert, W. and D. Dressler (1968) "DNA Replication: the Rolling Circle model." *Cold Spring Harbor Symp. Quant. Biol.*, 33: 473～478.
6. Kornberg, A. (1984) "DNA Replication." *Trends Biochem. Sci.* 9: 122～124.
7. Kornberg, A. (1988) "DNA Replication." *Biochim. Biophys. Acta* 951: 235～239.
8. McHenry, C. S. (1988) "DNA Polymerase Ⅲ Holoenzyme of *Escherichia coli*." *Ann. Rev. Biochem.* 57: 519～550.
9. McHenry, C. et al. (1984) "DNA Polymerase Ⅲ Holoenzyme of *E. coli*: an Asymetric Dimeric Replicative Complex Containing Distinguishable Leading and Lagging Strand Polymerases." In *Proteins Involved in DNA Replication* (edited by Hübscher, U. and S. Spadari, Plenum Press), pp. 315～319.
10. Kornberg, A. (1981) "DNA Polymerase – a perspective." *The Enzymes*, XIV: 3～13.
11. Maki, S. and Kornberg, A. (1988) "DNA Polymerase Ⅲ Holoenzyme of *Escherichia coli*. 1. Purification and Distinctive Functions of Subunits $\tau$ and $\gamma$, the dnaZX gene products." *J. Biol. Chem.* 263: 6547～6554.

12. Maki, S. and Kornberg, A. (1988) "DNA Polymerase III Holoenzyme of *Escherichia coli*. II. A Novel Complex Including the γ Subunit Essential for Processive Synthesis." *J. Biol. Chem.* **263**: 6555～6560.
13. Maki, S. and Kornberg, A. (1988) "DNA Polymerase III Holoenzyme of *Escherichia coli*. III Distinctive Processive Polymerases Reconstituted from purified Subunits." *J. Biol. Chem.* **263**: 6561～6569.
14. Maki, H. et al. (1988) "DNA Polymerase III Holoenzyme of *Escherichia coli*. IV. The Holoenzyme Is an Asymmetric Dimer with Twin Active Sites." *J. Biol. Chem.* **263**: 6570～6578.
15. Lehman, R. (1981) "DNA Polymerase I of E. coli." *The Enzymes*, **XIV**: 15～37.
16. Gefter, M. L. (1981) "DNA Unwinding Enzymes." *The Enzymes*, **XIV**: 367～372.
17. Gellert, M. (1981) "DNA Gyrase and Other Type II Topoisomerases." *The Enzymes*, **XIV**: 345～366.
18. Reichard, P. (1988) "Interactions Between Deoxyribonucleotide and DNA Synthesis." *Ann. Rev. Biochem.* **57**: 349～374.
19. Scott, J. and A. Kornberg (1978) "Purification of the Rep Protein of *E. coli*. An Enzyme Which Separates Duplex Strands in Adenovirus Replication." *J. Biol. Chem.*, **253**: 3292～3304.
20. Ogawa, T. and T. Okazaki (1980) "Discontinuous DNA Replication." *Ann. Rev. Biochem*, **49**: 421～428.
21. Benz, Jr. E. W. et al. (1982) "Priming Enzymes." *The Enzymes*, **XV**: 155～182.
22. Arai, K. and A. Kornberg (1981) "Unique Primed Start of Phage φX174 DNA Replication and Mobility of the Primosome in a Direction Opposite Chain Synthesis." *Proc. Nat. Acad. Sci.*, **78**: 69～73.
23. Ogawa, T. et al. (1985) "Initiation of Enzymatic Replication at the Origin of the *E. coli* Chromosome, Contributions of RNA Polymerase and Primase." *Proc. Nat. Acad. Sci.*, **82**: 3562～3566.
24. Van der Ende, A. T. et al. (1985) "Initiation of Enzymatic Replication at the Origin of the *E. coli* Chromosome: Primase as the Sole Priming Enzyme." *Proc. Nat. Acad. Sci.* **82**: 3954～3958.
25. Wechsler, J. A. and J. D. Gross (1971) "*E. coli* Mutants Temperature-sensitive for DNA Synthesis." *Mol. Gen. Genetics* **113**: 273～284.
26. Wechsler, J. A. (1978) "The Genetics of E. coli Replication." in *DNA Sythesis: Present and Future* (edited by M. Kohiyama and I. Molineux, New York, Plenum Press), p. 9.
27. Garcia, G. M. et al. (1986) "The *E. coli dna Y* Gene Encodes an Arginine Transfer RNA." *Cell* **45**: 453～459.
28. Zyskind, J.W. et al. (1986) "The Bacterial Origin of Replication, *oriC*." *Cell* **46**: 489～490.
29. Clayton, D. A. (1982) "Replication of Animal Mitochrondrial DNA." *Cell* **28**: 693～705.
30. Rokeach, L. A. et al. (1986) "RNA Terminating within the *E. coli* Origin of Replication, Stringent Regulation and Control by Dna A Protein." *Cell* **46**: 763～771.
31. Baker, T. A. et al. (1986) "Extensive Unwinding of the Plasmid Template During Staged Enzymatic Initiation of DNA Replications from the Origin of the *E. coli* Chromosome." *Cell* **45**: 53～64.
32. Chase, J. W. et al. (1986) "Single-stranded DNA Binding Proteins Required for DNA Replication." *Ann. Rev. Biochem.* **55**: 103～136.
33. Brill, S. J. et al. (1987) "Need for DNA Topoisomerase Activity as a Swivel for DNA Replication and for Transcription of Ribosomal DNA." *Natuer* **326**: 414～416.
34. Seufert, W. et al. (1987) "DnaA Protein Binding to the Plasmid Origin Region Can Substitute for Primosome Assembly During Replication of pBR322 in Vitro." *Cell* **48**: 73～78.
35. Patel, I. et al (1986) "A Replication Origin Is Turned off by an Origin Silencer Sequence." *Cell*: **47** 785～792.
36. Baker, T. et al. (1987) " Helicase Action of Dna B Protein During Replication from the *E. coli* Chromosome Origin in Vitro." *J. Biol. Chem.* **262**: 6877～6885.
37. Biswas, S. B. and E. E. Biswas (1987) "Regulation of DnaB Function in DNA Replication in *E. coli* DnaC and λ P Gene Products." *J. Biol. Chem.* **262**: 7831～7838.
38. Wickner, S.H. and D. K. Chattoraj (1987) "Replication of Mini-P1 Plasmid DNA in Vitro Requires Two Initiation Proteins, Encoded by the *repA* Gene of Phage P1 and the *dnaA* Gene of *E. coli*." *Proc. Natl. Acad. Sci.* **84**: 3668～3672.
39. Bramhill, D. and Kornberg, A. (1988) "Duplex Opening by Dna A Protein at Novel Sequences in Initiation of Replication at the Origin of the *E. coli* Chromosome." *Cell* **52**: 743～755.
40. Wahle, E. et al. (1989) "The DnaB-DnaC Replication Protein Complex of *Escherichia coli*. I. Formation and Properties." *J. Biol. Chem.* **264**: 2463～2468.
41. Wahle, E. et al. (1989) "The DnaB-DnaC Replication Protein Complex of *Escherichia coli*. II. Role of the

Complex in Mobilizing DnaB Functions." *J. Biol. Chem*. 264: 2469～2475.

42. Baker, T. A. etal.(1988)"Enzymatic Mechanism of Initiation of Replication from the Origin of the *Escherichia coli* Chromosome. "*Cancer Cells* 6: 19～24.

43. Schnos, M. et al. (1988)"Initiation Protein Induced Helix Destabilization at the λ Origin: A Prepriming Step in DNA Replication", *Cell* 52: 385～395.

44. Tsurimoto, T. and K. Matsubara(1982)"Replication of Bacteriophage λ DNA."*Cold Spring Harbor Symp. Quant. Biol*. 47: 681～691.

45. Fuller, C. W. et al. (1982) "Mechanisms for the Initiation of Bacteriophage T7 DNA Replication." *Cold Spring Harbor Symp. Quant. Biol*. 47: 669～677.

46. Nakai, H. and Richardson, C. C. (1988) "Leading and Lagging Strand Synthesis at the Replication Fork of Bacteriophage T7." *J. Biol. Chem*. 263: 9818～9830.

47. Nakai, H. and Richardson, C. C. (1988) "The Effect of the T7 and Escherichia coli DNA-binding Proteins at the Replication Fork of Bacteriaphage T7." *J. Biol. Chem*. 263: 9831～9839.

48. Van Mansfeld, A. D. M. et al.(1984)"Gene A Protein of Bacteriophage φX174 Is a Highly Specific Single-strand Nuclease and Binds Via a Tyrosyl Residue to DNA after Cleavage." in *Proteins Involved In DNA Replication*(edited by Ulrichs Hübscher and Silvio Spadari,Plenum Press) ,pp. 221～240.

49. Aoyama,A. et al.(1986)"Synthesis of Bacteriophage φX174 in Vitro:Mechanism of Switch from DNA Replication to DNA packaging." *Cell*, 47: 99～106.

50. McMacken, R. et al. (1978) "Priming of DNA Synthesis on Viral Single-stranded DNA in Vitro." in *The Single-Stranded DNA Phages* (edited by D. T. Denhardt et al. , Cold Spring hardor Laboratory), pp. 273～285.

51. Wickner, S. (1978) "Conversion of Phage Single-stranded DNA to Duplex DNA in Vitro." in *The Single-Stranded DNA Phages* (edited by D. T. Denhardt et al., Cold Spring Harbor Laboratory), pp. 255～271.

52. Kornberg, A. (1984) Enzyme Studies of Replication of the *Escherichia coli* Chromosome. in *Proteins Involved In DNA Replication* (edited by Ulrich Hübscher and Silvio Spadari, Plenum Press), pp. 3～16.

53. Kornberg, A.(1982)*1982 Supplement to DNA Replication*. (Freeman Press), pp. 124～127.

54. Watson, J. D. (1972)"Origin of Concatemeric T7 DNA." *Nature New Biol*. 239: 197～201.

55. Sussenbach, J.S. et al. (1972)"Linear Intermediates in the Replication of Adenovirus DNA" *Nature New Biol*. 239: 47～49.

56. Stillman, B. W. and F. Tamanoi (1982) "Adenoviral DNA Replication: DNA Sequences and Enzymes Required for Initiation in Vitro." *Cold Spring Harbor Symp. Quant. Biol*. 47: 741～750.

57. Stillman, B. W. et al. (1986) "DNA-protein Interactions at the Replication Origins of Adenovirus and SV40." in *DNA Tumor Viruses,Control of Gene Expression and Replication*(edited by M. Botchan et al., Cold Spring Harbor Laboratory) pp. 453～463.

58. Kenny, M. K. et al. (1988)"Initiation of Adenovirus DNA Replication I. Mechanism of Action of a Host Protein Required for Replication of Adenovirus DNA Templates Devoid of the Terminal Protein." *J. Biol. Chem*. 263: 9801～9808.

59. Kenny, M. K. and Hurwitz, J. (1988)"Initiation of Adenovirus DNA Replication II. Structural Requirements Using Synthetic Oligonucleotide Adenovirus Templates." *J. Biol. Chem*. 263: 9809～9817.

60. Sogo, J. M. et al. (1982)"Structure of Protein-containing Replicative Intermediates of *bacillus subtilis* Phage φ29 DNA ."*Virology* 116: 1～18.

61. Watabe, K. et al. (1984) "Replication of Bacteriophage φ29 DNA in Vitro:the Roles of Terminal Protein and DNA Polymerase." *Proc. Nat. Acad. Sci*. 81: 5374～5378.

62. Campbell, J. L. (1986) "Eukaryotic DNA Replication." *Ann. Rev. Biochem*. 55: 733～771.

63. Huberman, J. A. (1987) "Eukaryotic DNA Replication: A Complex Picture Partially Clarified." *Cell* 48: 7～8.

64. Fanning,E. et al.(1984)"How Does SV40 T Antigen Control Initiation of Viral DNA Replication?" in *Proteins Involved In DNA Replication* (edited by Ulrich Hübscher and Silvio Spadari,Plenum Press), pp.241～248.

65. Hay, R. T. and M. L. DePamphilis(1982)"Initiation of SV40 DNA Replication in Vitro: Location and Structure of 5'ends of DNA Synthesized in the *ori* Region."*Cell* 28:767～779.

66. DePamphilis, M. L. (1988) "Transcriptional Elements as Components of Eukaryotic Origins of DNA Replica-

tion." *Cell* 52: 635～638.

67. Mastrangelo, I. A. (1989) "ATP-Dependent Assembly of Double Hexamers of SV40 T Antigen at the Viral Origin of DNA Replication. Nature 338: 658～662.
68. Ryder, K. et al. (1986)"An Altered DNA Conformation in Origin Region I Is a Determinant for the Binding of SV40 Large T Antigen."*Cell* 44 :719～725.
69. Jones, K. A. et al. (1987)"A Cellular DNA-binding Protein that Activates Eukaryotic Transcription and DNA Replication." *Cell* 48: 79～89.
70. Prelich, G. et al. (1987) "The Cell-cycle Regulated Proliferating Cell Nuclear Antigen Is Required for SV40 DNA Replication in Vitro." *Nature* 326: 471～475.
71. Bravo, R. et al. (1987)"Cyclin/PCNA Is the Auxiliary Protein of DNA Polymerase-δ." *Nature* 326: 515～517.
72. Prelich, G. et al.(1987)"Functional Identity of Proliferating Cell Nuclear Antigen and a DNA Polymerase-δ Auxiliary Protein." *Nature* 326: 517～520.
73. Shampay, J. et al. (1984) "DNA Sequences of Telomeres Maintained in Yeast." *Nature* 310: 154～157.
74. Gottschling, D. E. and V. A. Zakian (1986) "Telomere Proteins: Specific Recognition and Protection of the Natural Termini of *oxytricha* Macronuclear DNA." *Cell* 47: 195～205.
75. Pülm, W. et al.(1984)"Chromatin Structure and Replication." in *Proteins Involved In DNA Replication* (edited by Ulrich Hübscher and Silvio Spadari, Plenum Press), pp. 127～141.
76. Stillman, B. (1986) "Chromatin Assembly During SV40 DNA Replication in Vitro." *Cell* 45: 555～565.
77. Greider, C. W. and Blackburn, E. H. (1987) "The Telomere Terminal Transferase of *Tetrahymena* Is a Ribonucleoprotein Enzyme with Two Kinds of Primer Specificity." *Cell* 51: 887～898.
78. Shippen-Lentz, D. and Blackburn, E. H. (1990) "Functional Evidence for an RNA Template in Telomerase." *Science* 247: 546～552.
79. Kornberg, A. (1980) "Regulation of Replication." in *DNA replication* (*Freeman Press*) *Chap.* 13.
80. Masukata, H. and J. Tomizawa (1986)"Control of primer Formation of ColE1 Plasmid Replication: Conformational Change of the Primer Transcript." *Cell* 44: 125～136.
81. Tomizawa, J. (1986) "Control of ColE1 Plasmid Replication: Binding of RNA I to RNA II and Inhibition of Primer Formation." *Cell* 47: 89～97.
82. Dasgupta, S. et al. (1987) "Multiple Mechanisms for Initiation of ColE1 DNA Replication: DNA Synthesis in the Presence and Absence of Ribonuclease H." *Cell* 51: 1113～1122.
83. Patel, I. and Bastia, D. (1987) "A Replication Initiator Protein Enhances the Rate of Hybrid Formation between a Silencer RNA and an Activator RNA." *Cell* 51: 455～462.

# 第四章　以修复作用为中心的DNA的安全保障体系

对于生命状态的存在和延续来说，DNA 分子要求保持高度的精确性和完整性，细胞中没有哪一种分子可以和它相比。在长期的进化中，生物体不仅演化出能纠正偶然的复制错误的系统，如 DNA 聚合酶的 3′→5′ 的校对功能，尿嘧啶 -N- 糖基酶修复系统；错配修复系统；而且还存在着能修复环境因素(如射线)和体内化学物质造成的DNA 分子的损伤的系统，如光复活修复系统、切除修复系统、重组修复系统、SOS 修复系统。为了对付外源 DNA 的入侵，有些生物还演化出限制修饰系统。从广义上来说，DNA 聚合酶的校对功能也可以算作修复系统；然而它是 DNA 聚合酶所具备的性质，在 DNA 聚合过程中发生，错误的碱基并不存在于DNA 链的内部，而只是瞬时存在于链的生长点上，因而仍然属于复制的范畴。而尿嘧啶 -N- 糖基酶修复系统和错配修复系统所要修复的错误碱基已经存在于新生 DNA 链的内部，因而本质上是属于修复范畴的；然而这两种修复系统主要作用于复制过程中和复制完成后的较短时间之内，因此编者将这两种修复系统归成一类，称之为复制修复(replicational repair)。由限制-修饰系统所引起的碱基甲基化也可以看作是复制后的加工(postreplicational processing)。但由于其生物学意义重大(虽然还不完全清楚)，所以限制与修饰也作为本章的重要内容。

## 第一节　复制修复[1~5]

### 一、尿嘧啶糖基酶系统

上一章中我们已经讲过，生物体内虽然有 dUTPase，但仍然会有少数 dUTP 分子逃遁而渗入 DNA 链中。由于 U 能和 A 发生氢键结合，PolⅢ 的校对功能无法识别它。另一方面，在细胞内由于胞嘧啶的自发的脱氨氧化而生成尿嘧啶。不管哪种来源的尿嘧啶都是通过尿嘧啶 -N- 糖基酶系统修复的。这一系统包括尿嘧啶 -N- 糖基酶，AP 内切核酸酶，PolI，DNA 连接酶。其修复过程示于图 4-1。

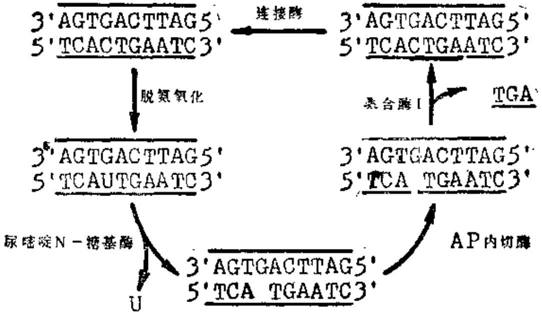



图 4-1　尿嘧啶 -N- 糖基酶系统的修复过程。本图所示为胞嘧啶脱氨氧化而生成的尿嘧啶的修复。复制时渗入的尿嘧啶的修复机制亦与此类似。

## 二、错配修复系统(mismatch repair system)

DNA 聚合酶偶尔能催化不能与模板形成氢键的错误碱基的渗入。这种复制错误通常由 DNA 聚合酶 3′→5′ 的校对功能立即进行纠正，然后才开始下一个核苷酸的聚合反应。然而在某些特殊条件下，DNA 聚合酶将极少数的错误碱基遗留在 DNA 链上而没有进行纠正。据估计，这种错误的频率为 $10^{-8}$ (即 $10^{8}$ 个碱基中有一个碱基是不配对的错误碱基)。然而，人们实际测量到的突变频率为 $10^{-10}$ 或 $10^{-11}$。由于 DNA 的完整性和精确性是生命的根本所在，因而细胞中演化出另一个修复系统，叫做错配修复系统，给予第二次纠正错误的机会。

然而，当新生的 DNA 链与模板链形成双螺旋结构(除不配对的部分外)，错配系统如何识别不配对部分的碱基哪一个应该被纠正呢？原来细胞中的甲基化作用与这种识别有关。在 DNA 中，天然的甲基化碱基有两种，一种是 $N^6$- 甲基腺嘌呤，一种是 5- 甲基胞嘧啶。胞嘧啶的甲基化似乎与此无关，不在这里介绍。腺嘌呤的甲基化则是错配修复系统的识别标志。这一标志具有碱基顺序特异性，$N^6$- 甲基腺嘌呤一般都包含在 GA*TC 这样的顺序中(A* 即

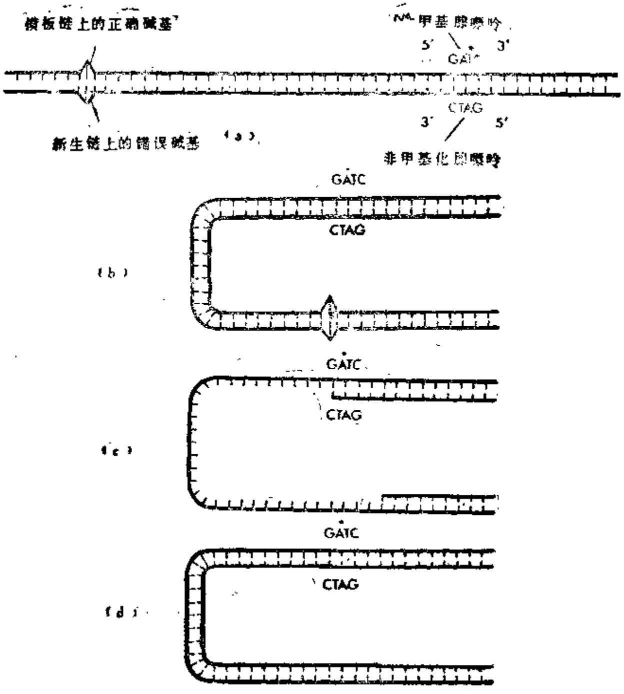



图 4-2 错配修复过程示意图。

(a) 模板链和新生链上不配对碱基及 GATC 序列的甲基化状况；(b) 错配矫正酶结合于未甲基化的 GATC 及同一条链上的错配碱基；(c) 错配矫正酶切除一段包含错配碱基的 DNA；(d) DNA 聚合酶进行缺口充填

$N^6$-甲基腺嘌呤);沿着新生 DNA 链,有一个甲基化的梯度,靠近复制叉处甲基化程度最小,亲本链上的甲基化程度高而且均一。根据这些特征,错配修复系统就可以识别出模板链和新生链,从而纠正新生链上的不配对碱基。

甲基化作用是甲基化酶所催化的。具体地说,$N^6$-甲基腺嘌呤是由 dam 基因所编码的 DNA 腺嘌呤甲基化酶催化所产生的。其甲基供体是 S-腺苷基 L-甲硫氨酸,迄今尚未见到其他甲基供体能满足这一反应的要求。错配修复系统主要包括一个能识别新生链上错配碱基和未甲基化的GATC序列的内切核酸酶,称为错配矫正酶(mismatch correction enzyme)。该酶由基因 *mutH*,*mutL* 和 *mutS* 所编码。此外,该系统还需要 DNA 聚合酶 I 和DNA 连接酶。错配修复过程示于图 4-2。

错配修复系统对于去除渗入 DNA 的碱基类似物也很重要。有些碱基类似物能够与模板上的碱基配对生成氢键而不能为 DNA 聚合酶的校对功能所识别;但是这种碱基类似物常有烯醇式和酮式的转变,因而在下一轮复制时则会引起突变。错配修复能够在下一轮复制之前将碱基类似物去除。此外,错配修复系统在基因转换过程中亦有重要作用,这种作用将在第十一章中讲授。

## 第二节 损伤修复[16~19]

生物体内外许多因素都能造成 DNA 分子结构的损伤,如电离辐射,紫外线,烷化剂,氧化剂等。一种因素可能造成多种类型的损伤;一种类型的损伤也可能来自不同种因素的作用。DNA 分子的损伤类型很多,其中以嘧啶二聚体,特别是胸腺嘧啶二聚体的形成和修复机制研究得最为清楚,也是本节阐述的重点。其他损伤类型及其修复机制只作简要的叙述。

### 一、胸腺嘧啶二聚体的产生

碱基在各种化学和物理因素如紫外线、电离辐射等的作用下,产生了碱基的衍生物,最常见的为胸腺嘧啶二聚体(图 4-3)。由于相邻的胸腺嘧啶产生二聚体,两个碱基平面被环丁基所扭转,引起双螺旋构型的局部变化,同时氢键结合力也显著减弱。这样,当有胸腺嘧啶二聚体的 DNA 作为模板进行复制时,Pol Ⅲ 将两个腺嘌呤核苷酸加上去,但由于不能很好地形成氢键,然后又由 3′→5′ 校对功能而将之水解。这样的事件反复发生,因而产生了一个空耗的过程,即大量的 dATP 被分解,而 DNA 复制则毫无进展。由于蛋白质仍在不断合成,而DNA不

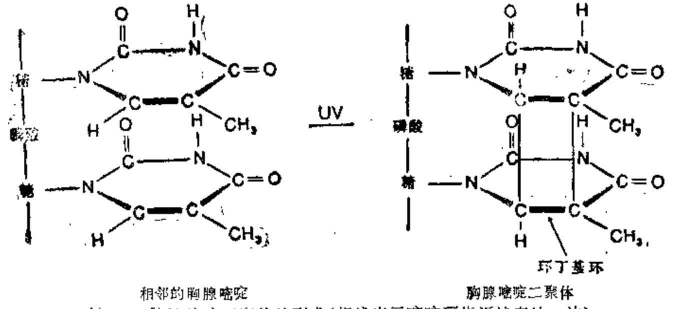



图 4-3 胸腺嘧啶二聚体的形成(粗线表示嘧啶环靠近读者的一边)

能复制，细胞也就不能分裂。这样就出现细丝状的所谓蛇形细胞，最后导致细胞死亡。

### 二、胸腺嘧啶二聚体修复的生物学指征

以紫外线照射为例，可以在照射的不同时刻取样，用菌落的形成能力为指标，将存活的百分比与照射时间（或剂量）作成曲线，称为存活曲线(survival curve)。见图 4-4 和 4-5。

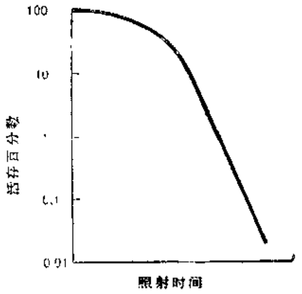


图 4-4　细菌的活存曲线

修复有多种类型，常见的主要有光修复和暗修复两类。光修复又叫光复活作用，就是细菌在紫外线照射后立即用可见光照射，这样可以显著地增加细菌的活存率，而且细菌的活存率随着可见光的剂量而增加(图 (4-5a、b)。如果将紫外线照射过的细菌在黑暗中置于非营养性的温暖的缓冲液中，也可以显著地增加细菌的活存率(图4-5c)，这种效应叫做液体保持恢复(liguid-holding recovery)，其实质是在这样的条件下有利于各种暗修复的进行。

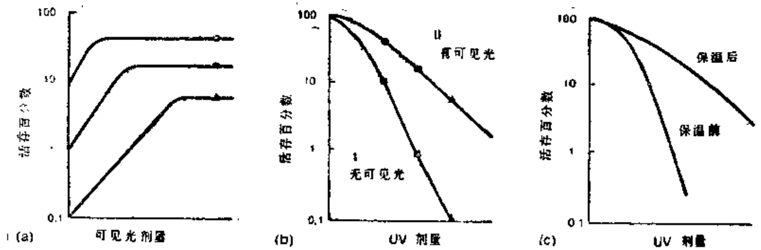


图 4-5　修复的类型

(a) 以不同剂量的紫外线照射细菌，然后用可见光照射，三种样品的活存率均为可见光剂量的函数
(b) 有无可见光照射的比较
　(Ⅰ) 没有可见光，曲线上三点取自图 a 的 y 轴上三点
　(Ⅱ) 有可见光，曲线上三点取自图 a 的高平部分的数值。
(c) 液体保持恢复效应

### 三、胸腺嘧啶二聚体修复的分子生物学机制

对于胸腺嘧啶二聚体，主要有五种修复途径：光复活、切除修复、重组修复、SOS 修复以及二聚体糖基酶修复机制。第一种为光修复，其余四种均为暗修复。光修复机制中利用光能来切除二聚体之间的 C–C 键，而暗修复中所需的能量来自 ATP 的水解。就目前所知，光修复机制只作用于紫外线照射所形成的产物，而暗修复除了可以作用于胸腺嘧啶二聚体外，还可以修复其他类型的损伤。*E.coli* 中，光修复过程只需要一个酶即光复活酶 (photoreactivating enzyme)，由基因 *phr* 编码；而暗修复作用则涉及到许多基因产物(表 4-1)。

1. 光复活(photoreactivation)　光复活作用是在可见光(300～600nm)的活化之下由光复活酶催化胸腺嘧啶二聚体分解成为单体的过程。几乎所有的生物细胞中都已发现光复活酶(简称 PR 酶)。光复活过程并不是 PR 酶吸收可见光，也不是 PR 酶与一个吸收可见光的化

表 4-1 *E.coli* 中与暗修复有关的基因产物

| 基　因 | 细胞中该基因突变的效应 | 基　因　产　物 | 功　　能 |
|---|---|---|---|
| *uvrA* | 对紫外线敏感 | 修复内切酶的 ATPase 亚基 | 除去胸腺嘧啶二聚体 |
| *uvrB* | 对紫外线敏感 | 修复内切酶的亚基 | |
| *uvrC* | 对紫外线敏感 | 修复内切酶的亚基 | |
| *uvrE* | 对紫外线敏感 | 未　知 | 未　知 |
| *recA* | 1. 重组缺陷<br>2. DNA 损伤后不能诱发修复途径 | recA 蛋白，40 KDa | 1. 重组及重组修复所需的 DNA 链交换活性<br>2. 诱导 SOS 修复系统所需的蛋白酶活性 |
| *recB* | 重组缺陷 | 外切核酸酶 V 的亚基 | 重组及重组修复所需 |
| *recC* | 重组缺陷 | 外切核酸酶 V 的亚基 | |
| *recD* | 重组缺陷 | 外切核酸酶 V 的亚基 | |
| *sbcB* | 抑制 *recBCD* 突变 | 外切核酸酶 I | 未　知 |
| *recF*<br>*recJ*<br>*recK* | 重组修复缺陷；在 *recBCD*、*sbcB* 突变体中重组缺陷 | 未　知 | 未　知 |
| *recE* | 重组缺陷 | 外切核酸酶 Ⅷ | 未　知 |
| *umuC* | 不能产生 SOS 修复的突变效应 | 未　知 | 未　知 |
| *lon* | UV 敏感，中隔缺失 | 结合于 DNA 的 ATP 依赖性的蛋白酶 | 控制荚膜多糖合成的基因 |
| *mutH*, *mutL*, *mutS* 等 *mut* 基因 | 增加突变率 | 错配矫正酶亚基 | 错配修复系统所需 |
| *dam* | UV 敏感，增加突变率 | DNA 腺嘌呤甲基化酶 | 错配修复系统所需 |
| *lexA* | 使许多修复系统基因，特别是SOS 系统的基因失去调节 | 阻遏蛋白，40KDa | 控制许多修复系统基因的表达。 |

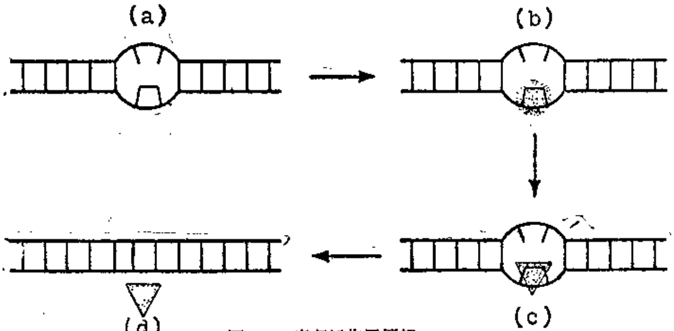



图 4-6　光复活作用图解

(a) 由于胸腺嘧啶二聚体而引起 DNA 变形，　(b) PR 酶～DNA 复合物的形成，
(c) 可见光的吸收和 PR 酶的活化，　(d) DNA 双链复原和 PR 酶的释放。

合物相结合;而是 PR 酶先与 DNA 链上的胸腺嘧啶二聚体结合成复合物;这种复合物以某种方式吸收可见光并利用光能切断胸腺嘧啶二聚体之间的两个 C-C 键。胸腺嘧啶二聚体变为单体,PR 酶就从 DNA 上解离下来(图 4-6)。其具体的光化学机制还不清楚,在紫外线照射下,细胞中还可以形成胞嘧啶二聚体以及胞嘧啶-胸腺嘧啶二聚体,光复活作用同样能使之单体化。

2. 切除修复　切除修复是一种多步骤的酶反应过程。过去,人们常常将这四个步骤归纳为"切—补—切—封"四个步骤。第一步为切开,一种修复内切酶能够识别胸腺嘧啶二聚体所引起的 DNA 双螺旋结构的变形,并在二聚体前头的糖-磷酸骨架上作一切口,切口一边是 5′-P,另一边是 3′-OH。第二步是由 PolI 在 3′-OH 末端聚合一条新的 DNA 链并同时置换掉大约 20 个核苷酸的 DNA 片断(其中包括胸腺嘧啶二聚体)。第三步,被置换出来的片断由 PolI 的 5′→3′ 外切核酸酶活性和内切核酸酶活性切除。除了 PolI 之外,还有其他的外切核酸酶也能进行这一反应。如 DNA 聚合酶 Ⅱ 的 5′→3′ 外切酶活性和具有单链特异性的外切核酸酶Ⅶ。因此,这三个基因中任何一个发生突变,对去除胸腺嘧啶二聚体都没有显著的影响。但是根据双突变体的研究,人们普遍认为 DNA 聚合酶 I 是执行这项任务的主要的酶。第四步是 DNA 连接酶封合新合成的 DNA 片断和原有 DNA 链之间的一个切刻。某些噬菌体编码的修复内切酶确实是按上述方式作用的。

*E.coli* 的修复内切酶由三种亚基构成:UvrA、UvrB、UvrC。其中UvrA 具有 ATP 酶活性。UvrA 加上 UvrB 在离体情况下就有活性,而 UvrC 对于该内切酶达到最大活性是不可缺少的。其精细作用还不清楚。这三个基因中任何一个基因发生突变,都会增加细菌对于紫外线的敏感性。修复内切酶对于胸腺嘧啶二聚体的识别原理是因为胸腺嘧啶二聚体引起了 DNA 双螺旋的变形。因而,Uvr 系统不但能修复胸腺嘧啶二聚体,而且能修复其他类型的 DNA 损伤,只要这些损伤能引起 DNA 双螺旋的变形,就能为修复内切酶所识别。

长期以来,人们以为*E.coli* 的修复内切酶也是采取在损伤的 5′ 端作一切口的方式进行修复的。通过基因克隆,人们分离到大量的 Uvr ABC 产物以供作用机制的研究。结果发现 Uvr ABC 在胸腺嘧啶二聚体的两侧都作一切口,切下包含损伤在内的 12 个核苷酸的单链片

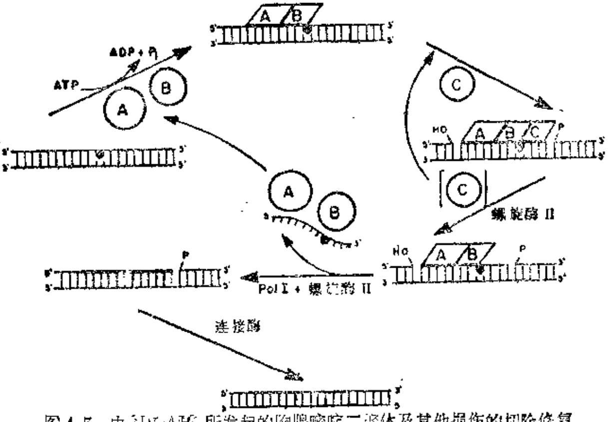



图 4-7　由 UvrABC 所发起的胸腺嘧啶二聚体及其他损伤的切除修复

断，留下的12个核苷酸的缺口由DNA聚合酶I进行充填；最后由DNA连接酶连接起来(图4-7)。显然，由于UvrABC能够修复不同类型的损伤，因此它并不能直接识别任何一种损伤部位；相反，它识别的是DNA某一部位缺少正常的形状。UvrABC首先结合于不含胸腺嘧啶二聚体的一侧，然后包裹到DNA双螺旋的另一侧。当它“触摸”到非正常形状时，立即触发了它的核酸内切酶活性，在双螺旋的不含嘧啶二聚体的一侧，在一圈多一点的两个位点上切下包含二聚体的12个核苷酸单链DNA片断。

根据切除的DNA片断的长短，可以分为“短补丁修复”(short-patch repair)和“长补丁修复”(long-patch repair)，大约99%的切除修复为短补丁修复，切除的DNA一般不超过30个核苷酸，其中包括一次切除的12个核苷酸。剩下的1%为长补丁修复，切除的DNA大多在1 500核苷酸左右，少数可达9 000核苷酸以上。参加这两种修复过程的酶大多数是相同的，在长补丁修复过程中可能还有别的因子参与，然而人们对这种因子还不清楚。但人们相信，短补丁修复是细菌细胞的组成型功能，而长补丁修复是DNA损伤所诱导出来的功能。

许多生物中都有修复内切酶。大肠杆菌、微球菌、酵母菌以及许多哺乳动物的细胞中的Uvr内切酶都已作过详细的研究。细菌中的修复内切酶还能修复噬菌体的紫外线损伤。某些噬菌体甚至有它自已的修复内切酶系统，如大肠杆菌噬菌体T4。

患有干皮性色素沉着(xeroderma pigmentosum)的病人就是由于为紫外线内切酶系统编码的基因发生突变而缺乏切除修复的能力。这种病人在阳光下就会发生皮肤损伤甚至出现几种皮肤癌。

3．重组修复　上面我们已经讲过了Uvr系统，当然，Uvr系统中任何一个基因突变体对于紫外线都要比野生型细菌敏感得多，其中*uvrA*⁻尤甚(图4-8)。然而，对于*uvrA*⁻，平均使一个细胞死亡的紫外线剂量能在细胞内产生300个胸腺嘧啶二聚体，实验表明这些二聚体都没有被切除；而我们又假定(事实也证明了)没有被修复的一个二聚体就足以使细胞死亡。因而，在正常细胞中，除了Uvr系统外，必定还有其他修复系统以不同的途径消除胸腺嘧啶二聚体的后果。人们发现，如果正常的遗传重组所必需的基因*recA*发生突变，则对于紫外线更加敏感(图4-8)。因此*recA*基因必定是作为另一种修复系统的成员而起作用；而且在*recA*⁻中，二聚体的切除能正常地进行，所以*recA*⁻所参与的修复机制与切除修复完全不同。从图4-8可以看出，*recA*⁻对于紫外线比*uvrA*⁻更为敏感，表明Rec修复系统比切除修复系统更为有效。通过双突变*uvrA*⁻*recA*⁻的研究更可以证明这两个相对独立的修复系统的存在，因为双突变体*uvrA*⁻*recA*⁻对紫外线的敏感程度比任何一个单突变体*uvrA*⁻或*recA*⁻都要高。在*uvrA*⁻*recA*⁻双突变体中，引起细胞死亡的胸腺嘧啶二聚体数大约为1。现已知道，Uvr系统负责切除大量的胸腺嘧啶二聚体，而Rec系统则负责消除那些没有被切除的二聚体可能造成的可怕的后果。

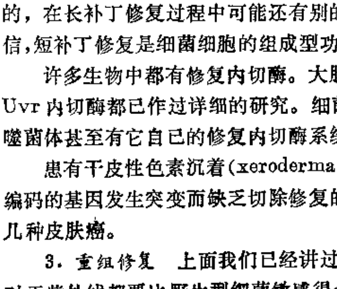



图4-8　野生型、*uvrA*⁻、*recA*⁻三种*E.coli*的活存曲线

前面我们已经讲过，胸腺嘧啶二聚体会引起PolⅢ在二聚体上进行反复的聚合和水解反应，使细胞耗费了大量的dATP前体而没有DNA的合成，最终导致细胞的死亡。然而在

$rec^+$ 细胞中(不管 $uvr^+$ 或 $uvr^-$)，DNA 合成的抑制作用一般不是永久性的；除非紫外线剂量很大,造成 Rec 系统本身的损伤。细胞中有两种途径可以使 DNA 复制继续进行，一种是所谓二聚体后起始(postdimer initiation)由重组修复系统负责;另一种是所谓超越二聚体合成(transdimer synthesis),由 SOS 修复系统负责。关于 SOS 修复系统,我们将在下一个修复机制中讲述。

重组修复系统中，DNA 复制到胸腺嘧啶二聚体时,大约暂停 5 秒钟左右，而在二聚体的后面(即顺流方向)又以一种未知的机制起始 DNA 的复制,这种起始很可能是不需要引发的。这样,在合成的子链上,就有许多大的缺口,模板链上有一个二聚体，子链上就有一个缺口(图 4-9)。这样的 DNA 分子不经修复，是不可能再复制下去的。因为再复制下去只能产生短的 DNA 片断。然而通过姐妹链交换的机制,就能产生一条完整的 DNA 链，以便作为下一轮复制的模板(图 4-10)。虽然原来的二聚体依然存在，但细胞却完成(勉强完成)这一轮复制。留下的二聚体可能被其他修复机制所去除;或者虽然没有去除，但随着细胞分裂的进行，这种损伤的 DNA 在细胞群体中逐渐地被“稀释”。

图 4-9 二聚体后起始的复制与子链中的缺口

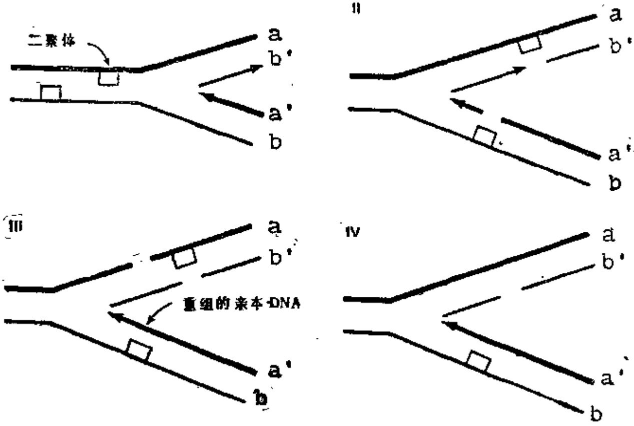



图 4-10 重组修复的姐妹链交换假说图解

(Ⅰ) DNA 分子的 a 链和 b 链上各有一个胸腺嘧啶二聚体并开始复制;
(Ⅱ) 通过二聚体后起始机制,复制出包含缺口的子链 a′ 和 b′;
(Ⅲ) a′,a 姐妹链发生交换，即从 a 链上切下 a′ 链上所缺少的一段，由 DNA 连接酶连接在 a′ 链上。在下一轮复制时，a′ 链成为完整的模板;
(Ⅳ) a 链上的缺口由 PolI 和 DNA 连接酶补好。b′ 链可以按同样的方式进行修复。
但是 a 链上的二聚体如果被其它机制所修复，那末只要由 DNA 聚合酶 I 补齐就行了

图 4-10 所示的重组修复机制仅系假说而已,其具体分子机制还不知道。当两条链上的两个二聚体处于非常靠近的位置时,则重组修复系统无法进行修复。

重组修复是非常重要的修复机制，它不必像切除修复那样需要等待很久(几个小时)细胞 DNA 才能复制，而是先复制再修复，因而重组修复又称为复制后修复(postreplicational repair)。重组修复还能修复某些不能为切除修复所去除的损伤。例如有些损伤并不引起 DNA 双螺旋的变形,但确实阻碍复制的进行。在许多细菌中都发现了重组修复，但是还不知道动物细胞中是否也有重组修复。重组修复所涉及的基因大多是细胞内正常的遗传重组所需的基因,但也有些基因的突变只影响其中一个过程,因此重组修复和正常的遗传重组并不是完全一致的。在这个系统中，最关键的基因是 *recA*。RecA 具有催化 DNA 分子之间的同源会联和交换单链的功能。这对于正常的遗传重组和重组修复都是必不可少的。这两种系统还需要 *recBCD* 基因，RecB，RecC和 RecD 是外切核酸酶 V 的亚基。该酶的活性还受 rec 系统其他组分的控制；在 *recA*⁻ 突变体中，该酶则不受约束，表现出鲁莽的行为即降解大量的 DNA。基因 *recF*,*J*,*K*,E 的性质和功能都还不清楚。

重组修复和切除修复两个系统也不是互相独立、完全无关的。在鉴定修复缺陷突变体时常常发现某一突变体与两个系统都有关联,可能某些酶(如外切酶、聚合酶等)同时参与了两种系统或者某种酶的功能可被另一种酶所代替。这里值得一提的是 RecA 这一特别的蛋白质，它不但有重组活性，而且还有单链 DNA 结合活性以及由此产生的蛋白酶活性。正是这种蛋白酶活性诱发了许多基因特别是修复系统基因的表达,其中包括切除修复系统,重组修复系统和下面要讲的 SOS 修复系统。

4. SOS 修复　SOS 修复是一种旁路系统，它允许新生的 DNA 链越过胸腺嘧啶二聚体而生长,其代价是保真度的极大降低,这是一个错误潜伏的过程。有时尽管合成了一条和亲本一样长的 DNA 链,但常常是没有功能的。其原则是:丧失某些信息而活存总比死亡好一些。SOS 修复的机制至今还没有透彻地理解,但据认为是 SOS 修复系统引起校对系统的松懈，以使聚合作用能够向前越过二聚体(超越二聚体合成),而不管二聚体处双螺旋结构的变形。

当 SOS 修复系统被活化以后，校对系统就大大地丧失了识别双螺旋结构变形的能力，从而结束了上面说过的空耗过程，而使复制又继续前进。两个腺嘌呤也加到与二聚体相对的位置上。关于 SOS 修复的一些早期的假说认为,在二聚体相对的位置上,可以是任何碱基,因而引起突变。但是新近的研究结果表明,错误的碱基可以出现在生长链上的任何位置,而不一定要在二聚体相对的位置上。由于校对功能的丧失，在新合成的链上有比正常情况下多得多的不配对碱基。尽管这些错配碱基可以被错配修复系统和切除修复系统纠正,但因数量太大,没有被纠正的错配碱基仍然很多。现在看来，SOS 修复系统的介入是紫外线诱发突变的主要原因。SOS 系统可能以某种方式对 PolⅢ 进行修饰，如改变某一个负责校对功能的亚基。基因 *umuC* 的产物可能是这个错误潜伏的修复系统的重要组分，因为 *umuC* 的突变使SOS 的诱变作用丧失。

SOS 修复(或许还有重组修复)与切除修复不同，切除修复系统的酶在正常细胞中就已存在;而 SOS 系统的酶只到细胞受到损伤时才出现。这可以从这样的一个实验得到证明：紫外线照射过的*E. coli*比没有照射过的 *E. coli* 更能支持紫外线照射过的 λ 噬菌体的生长(图 4-11)。这种现象,叫做 UV 复活(UV-reactivation)或者 W 复活(为了纪念这一现象的发现者 Jean Weigle)。这个现象显然牵涉到一种修复过程,而且只有当细胞是 *recA*⁺ 时才出现；然

而它并不需要 *uvr* 基因。现在认为，这种现象就是因为 SOS 系统在起作用；当紫外线照射细菌时，SOS 系统就被开启。

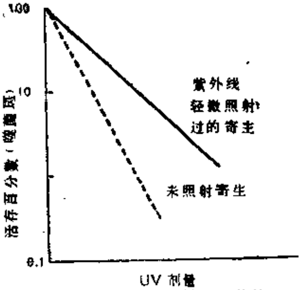


图 4-11 紫外线照射的噬菌体 λ 的 W复活现象

SOS 修复系统，牵涉到几个基因，其中最清楚的是 *recA* 基因和 *lexA*基因(即对紫外线抗性基因)。在正常细胞内，SOS 系统是关闭的。生物在长期进化过程中，已经有了优良的校对功能以在复制中保持极高的准确性，因此必须使 SOS 这个错误潜伏的复制过程关闭。这个作用是通过 *lexA* 基因产物而实现的。LexA蛋白是一种阻遏蛋白，结合于 SOS 系统中各基因的操作子(operator)上，使得这些基因没有活性，很少产生转录产物，从转录水平上控制了这些基因。再说 *recA* 基因，它的产物有三种主要的生物化学活性：一是重组活性；二是单链 DNA 结合活性；三是蛋白酶活性。在细胞内，DNA 合成正常进行时，RecA 蛋白是没有蛋白酶活性的。然而，当 DNA 合成受到阻碍时，一部分已经存在于细胞中的 RecA 蛋白就转变成有活性的蛋白酶。这种活化作用需要单链 DNA，很可能与 RecA 蛋白结合，因为 RecA 蛋白有单链 DNA 结合活性。RecA 蛋白酶活性仅作用于少数特定的蛋白质，LexA 蛋白就是其中之一。LexA 蛋白被水解成两个片断，就不再阻止 SOS 的基因的转录，因而就开启了 SOS 系统。当修复完成时，DNA 合成转入正常，RecA 蛋白又失去了蛋白酶活性，LexA 蛋白又得以控制 SOS 系统的基因的转录，从而使 SOS 系统处于关闭状态。SOS 修复过程示于图 4-12。

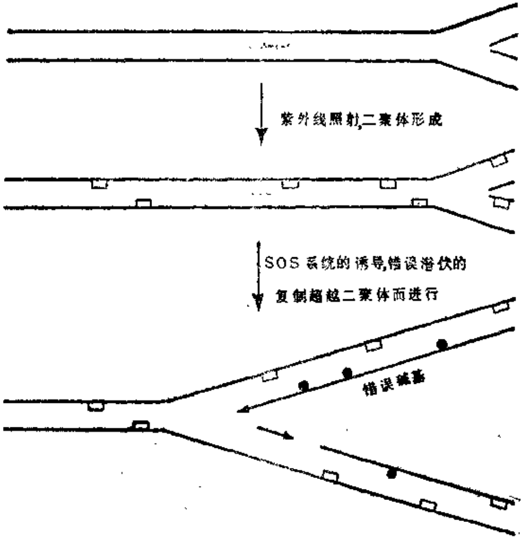


图 4-12 SOS 修复过程

SOS 修复只是 SOS 反应(SOS response)的一部分。细菌在 UV、丝裂霉素 C、烷化剂等作用下,造成了 DNA 的损伤。从而抑制了 DNA 的复制。在这种情况下,细胞就会产生一系列的表型变化,包括对损伤 DNA 的修复能力迅速增强,诱变率的提高,细胞分裂停止以及 λ 原噬菌体的诱导释放等。这些反应通称为 SOS 反应。不单造成 DNA 损伤的因素可以引起 SOS 反应,其他凡能抑制 DNA 复制的因素,如胸腺嘧啶饥饿,加入 DNA 复制的抑制剂以及某些重要的 *dna* 基因的突变等,均能触发 SOS 反应。

上面已经说到,RecA 在 SOS 反应中起着核心作用。*recA* 基因长 1 059 bp,包含 353 个氨基酸,分子量 40 KDa。关于 RecA 的蛋白酶活性的产生机制仍不清楚,只是体外试验时,RecA 需要在单链 DNA 存在时才表现出蛋白酶活性。很可能是由于 RecA 与单链 DNA 结合后产生分子内变构,产生蛋白酶活性。这并不排除在体内由于损伤 DNA 而产生的某种小分子中间物激活了 RecA 的机制。不管存在什么机制,这种激活作用非常之快,在 DNA 受伤后的几分钟内就会出现 SOS 反应。

RecA 的蛋白酶活性是一种特异的内肽酶,只作用于少数几种蛋白质的丙氨酸-甘氨酸之间的肽键,将底物一分为二。在 *E. coli* 中,除了 LexA 蛋白以外,还有 λ 噬菌体的阻遏蛋白。当 LexA 基因发生突变,使 Ala-Gly 变为 Ala-Asp 时,就不能诱发 SOS 反应。能被 RecA 蛋白酶活性水解的有限几种蛋白质除了 Ala-Gly 序列外,其他部位的氨基酸序列并无同源性;而且并非含有 Ala-Gly 的所有蛋白质都能成为 RecA 的底物。由此可见,底物的空间结构可能起重要作用。

LexA 蛋白是一种阻遏蛋白,分子量 22KDa,由 *lexA* 基因编码。在正常细胞内 LexA 蛋白非常稳定,控制着许多操纵元的表达,特别是各修复系统的基因。通过操纵元融合 (operon fusion)试验,已发现受 LexA 控制的基因有 *recA*, *lexA*, *uvrA*, *uvrB*, *umuC*, *himA* 等 17 个基因,它们统称为 *din* 基因(damage inducible genes),又称为 SOS 基因。这些基因中,有的基因只在 DNA 复制受到抑制时才表达;有时在正常细胞中就有低水平的表达,而在 DNA 受

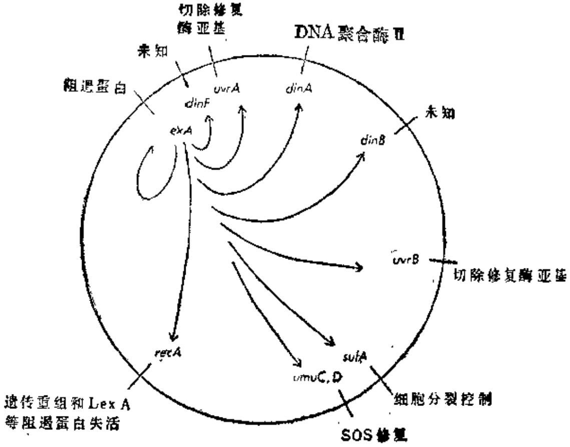



图 4-13 由 LexA 控制的 *E. coli* SOS 基因构成一个调控元。

到损伤时表达量急剧增加。例如，*uvrB* 基因，它是切除修复系统的成员之一，该基因有两个启动子，一个不受 LexA 的控制，另一个受 LexA 的控制。当 LexA 被 RecA 降解后，两个启动子均能发起 *uvrB* 基因的转录。由 LexA 所控制的 SOS 基因构成了 *E. coli* 中最大的调控元(图 4-13)。

所有 SOS 基因的操作子都含有 20bp 的 LexA 结合位点，称为 SOS 框(SOS box)。SOS 框的一致序列是 $\genfrac{}{}{0pt}{}{T}{A}$ACTGTATATNCATNCAGGA，它具有一定的对称性，其中底下带点的八个位置绝对保守。与大多数的操作子一样，SOS 框与各自的启动子序列相重叠。*lexA* 基因受其产物 LexA 蛋白的控制，称为自身负向控制，因为 *lexA* 基因具有前后相连的两个 SOS 框。在正常细胞内，由于 LexA 基因的自身负向控制而处于一种稳定的表达水平，其产物 LexA 蛋白足以占据细胞内所有 SOS 框，阻遏或部分阻遏各SOS基因的表达。当细胞的 DNA 复制受阻时，某种信号激活了 RecA 蛋白酶活性，水解了 LexA，(这一过程非常迅速)，使 SOS 系统的各个基因的表达水平迅速提高。特别是 RecA 基因表达增加极快。在大肠杆菌的对数生长期，每个细胞大约有 2 000～5 000 个 RecA 分子，而在引起 DNA 损伤的处理之后(如紫外线，丝裂霉素 C，萘啶酮酸等)，迅速增加到 150 000 个分子。这大概与 RecA 的多方面功能有关。值得一说的是 *lexA* 基因的表达亦大量增加，乍看起来有些费解。因为 SOS 系统的激活就是要水解掉 LexA。当然毫无疑问，在 SOS 系统的激活过程中，大量表达的 LexA 注定要被 RecA 降解掉。然而细胞一旦度过了 DNA 复制受阻的难关，则 RecA 的蛋白酶活性很快消失；而大量产生的 LexA 又立即占领所有 SOS 框，从而迅速关闭了 SOS 系统。因为 SOS 修复过程是一个错误潜伏的过程，所以细胞不到万不得已是不会启动这一系统的。这也就解释了为什么 SOS 反应可以迅速逆转。

当 RecA 的蛋白酶活性作用于 λ 噬菌体的 CI 阻遏蛋白时，就结束了 λ 噬菌体的溶原周期，使之进入裂解周期。这一作用并不是细胞的功能，而是 λ 噬菌体的本能反应，它了解到寄主自己已遇到了麻烦，还是“走为上计”——进入裂解周期，产生更多的后代噬菌体。

5. 嘧啶二聚体糖基酶修复系统　当 T4 噬菌体感染大肠杆菌细胞时，其基因 *denV* 的产物即为嘧啶二聚体糖基酶。该酶同时兼具 AP 内切酶活性。其修复机制与非标准碱基的修复机制基本相同。

上面所述的五种修复机制中，从能量来源方面来说，只有光复活利用光能，其余利用 ATP 水解所释放的能量。光复活，切除修复和二聚体糖基酶修复都是修复模板链，重组修复是形成一条新的模板链，SOS 修复是产生连续的子链。SOS 修复是唯一导致突变的修复，而其余的修复机制是将损伤的 DNA 恢复到损伤前的状态或者是产生与亲本相同的子代 DNA。

最后人们可能会有一个疑问，既然紫外线并不是大多数细胞经常碰到的因素，那么为什么生物细胞中有一系列修复紫外线损伤的修复系统呢？细胞的性质总是反映古代的状况。当生命从原始海洋中出现时，地球周围还没有臭氧层，高通量的紫外线源源不断到达地球表面，生物有机体就必须有能够修复紫外线损伤的机制，否则就会被淘汰。这些修复机制不但能修复紫外线损伤，而且能修复其他因素所造成的损伤，如切除修复机制就是如此。

### 四、其他损伤类型及其修复

1. 非标准碱基　在 DNA 分子中，除了尿嘧啶能够在 DNA 复制时渗入以外，胞嘧啶也能自发地脱氨氧化成为尿嘧啶。腺嘌呤也可以脱氨氧化生成次黄嘌呤。此外，还有些其他的

碱基衍生物会存在于 DNA 链中，例如 3- 甲基腺嘌呤，6- 氢 -5,6- 二羟胸腺嘧啶，2,4- 二氨基 -6- 羟 -5-N- 甲基甲酰亚氨嘧啶，嘧啶二聚体等。对于这些非标准碱基(图 4-14))，均已发现其相应的 DNA 糖基酶(glycosylases)。

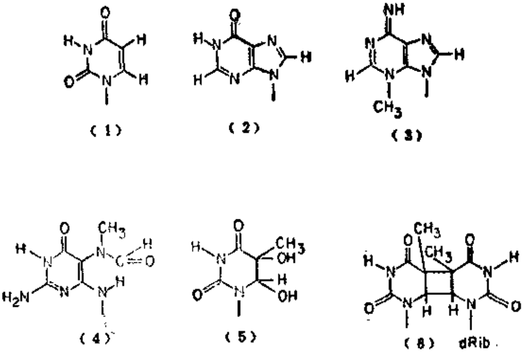



图 4-14　已发现相应糖基酶的若干非标准碱基

(1) 尿嘧啶；(2) 次黄嘌呤；(3) 3-甲基腺嘌呤；(4) 2，4-二氨基-6-羟-5-N-甲基甲酰亚氨嘧啶；(5) 6-氢-5,6-二羟胸腺嘧啶；(6) 胸腺嘧啶二聚体

DNA 糖基酶广泛存在于几乎所有的生物体内。根据它们是否具有内在的 AP 内切酶活性，可以将 DNA 糖基酶分为两类：第一类酶不具有 AP 内切酶活性，如尿嘧啶糖基酶，次黄嘌呤糖基酶，3- 甲基腺嘌呤糖基酶，甲酰亚氨嘧啶糖基酶等。第二类酶具有内在的 AP 内切酶活性，如嘧啶二聚体糖基酶，水合胸腺嘧啶糖基酶。

DNA 糖基酶修复系统对于生物体的生存十分重要。如果这些酶的基因发生突变，细菌生长状况很差。DNA 糖基酶的修复机制，和前面我们在复制性修复一节中讲的尿嘧啶糖基酶相同。但是，在第一类 DNA 糖基酶除去非标准碱基之后，除了 AP 内切酶途径之外，还可以通过 DNA 碱基插入酶进行进一步的修复(见下文所述)。

2. *碱基的丢失*　DNA 中的脱氧核糖，一方面使主链更为稳定，另一方面使其糖苷键的稳定性低于 RNA。DNA 的双螺旋结构极大地克服了这一欠缺，但是嘌呤碱基仍然容易自发水解。这一过程称为去嘌呤作用(depurination)。在生理条件下，自发的去嘌呤速度为$3\times10^{-11}$秒$^{-1}$，而自发的去嘧啶速度要低 20 倍。在实验室中可以用降低 pH 和提高温度的办法促进去嘌呤作用。某些烷化剂，如乙磺酸乙酯($CH_3CH_2SO_3CH_2CH_3$)，氮芥气之一甲基-二 (β- 氯乙基)氮$\left(CH_3-N\left\langle\begin{matrix}CH_2CH_2Cl\\CH_2CH_2Cl\end{matrix}\right.\right)$，能在嘌呤环(特别是鸟嘌呤)的第 7 位氮原子上接上烷基。这种烷化作用极大地减弱了第 9 位 N 原子上的糖苷键，因而发生去嘌呤作用。

细胞内有一种内切核酸酶叫做无嘌呤内切核酸酶，简称 AP 内切酶。它不但负责修复去嘌呤作用造成的损伤，而且也负责修复 DNA 中的尿嘧啶和次黄嘌呤等。因而这一修复过程是 N- 糖基酶修复系统的修复过程的一部分。AP 内切酶实际上应称为无碱基内切酶，它能特异地切断无碱基核糖的 5′-磷酸的 3′- 酯键。

细胞内还存在着一种能直接将适合的碱基插入 AP 位点的酶，称为碱基插入酶，简称插

入酶(insertase)。目前只发现一种嘌呤碱基插入酶，而没有发现嘧啶碱基插入酶。嘌呤碱基插入酶的专一性极强，只能作用于双链 DNA 上的 AP 位点，而不能作用于单链 DNA 上的 AP 位点；插入的嘌呤碱基严格遵循碱基互补的规则，即根据互补链上嘧啶碱基状况，在AP位点上插入 A 或 G；对于无嘧啶位点不起作用；嘌呤碱基的供体可以是自由的嘌呤碱基，也可以是嘌呤脱氧核苷，而不可以是非脱氧的嘌呤核苷。人的嘌呤碱基插入酶不能利用 dATP 和 dGTP，而 *E. coli* 的酶可以利用这两种脱氧核苷三磷酸。这种酶在形成糖苷键时无疑是需要能量的，但能量的来源至今仍然是个谜。至少是人类嘌呤碱基插入酶不需要任何富能辅因子。据信，DNA 双螺旋的碱基再配对和碱基堆集作用是自由能的来源之一。

碱基插入酶的生物学意义仍然不清楚。有人认为它为修复 AP 位点提供了一条更为迅捷的途径，这一途径当 AP 内切酶基因突变时就更为重要。还有人认为，碱基插入酶与 DNA 糖基酶相互配合，迅速实现某些 DNA 序列上的修饰碱基向非修饰碱基转变，可能在触发某些发育和调节机制中起重要作用。

3．烷基化损伤的修复　许多烷基化试剂(如亚硝基胺类和亚硝基胍类)都能将 DNA 中的嘌呤碱基特别是鸟嘌呤烷基化，除 N-7 和 N-3 烷基化外，还能在 O-6 位置上烷基化以及在磷酸骨架上出现烷基化。这后两种烷基化能够被一种叫做 $O^6$-甲基鸟嘌呤甲基转移酶($O^6$-methylguanine methyltransferase)所修复。该酶由 *ada* 基因所编码。该酶有 354 个氨基酸残基，分子量为 39KDa。Ada 蛋白不但能去除鸟嘌呤 $O^6$ 位上的甲基，而且可以去除甲基磷酸三酯上的甲基。被去除的两种甲基分别结合于该酶 $Cys_{321}$ 和 $Cys_{69}$ 这两个半胱氨酸残基上。然而，已经甲基化的酶却无法再转变为非甲基化的酶。这种现象被称为该酶的自杀行为。这样，从 DNA 分子上每除去一个甲基就要消耗一个新的酶分子。

人们还发现了大肠杆菌对于烷基化试剂有一种适应性反应(*ada*基因便由此得名)。例如，将大肠杆菌先与相对来说"无害"的低浓度亚硝基胍接触，而后则能抵抗浓度高出上百倍的这种突变剂。与此同时，细胞内 $O^6$-甲基鸟嘌呤甲基转移酶和烷基化 DNA 糖基酶含量急剧增加。而 *ada* 基因的突变则阻碍了这种诱导作用。可见 *ada* 基因产物不但作为修复烷基化的酶，而且作为自身基因和烷基化 DNA 糖基酶基因的诱导物。更确切地说，当 $O^6$-甲基鸟嘌呤甲基转移酶携带了甲基之后就变为诱导物。这一诱导物共诱导四个基因：*ada*, *alkA*, *alkB*, *aidB*，其中 *ada* 和 *alkB* 组成一个操纵元。*alkA* 基因编码烷基化 DNA 糖基酶(曾经叫做 3-甲基腺嘌呤-DNA 糖基酶 Ⅱ)，该酶能除去 3-甲基腺嘌呤，3-甲基鸟嘌呤，$O^2$-甲基胞嘧啶，$O^2$-甲基胸腺嘧啶等烷基化碱基。*alkB* 基因产物负责细胞毒素损伤的 DNA 的切除修复。*aidB* 基因的功能尚不清楚。上述四个基因在其 RNA 聚合酶结合位点的上游都有诱导物的结合位点，其一致序列为 AAANNAAAGCGCA。

4．链的断裂　许多因素能产生链的断裂，如过氧化物，巯基化合物，某些金属离子以及 DNAase 等。电离辐射具有强烈的链断裂作用，这是通过辐射粒子的直接作用和间接作用造成的。所谓间接作用就是辐射粒子在体内产生次级高能电子和自由基(如 $OH^{\cdot}$、$H^{\cdot}$、$HO_2^{\cdot}$ 等)然后作用于 DNA 而产生链的断裂。如果是一处双链同时断裂，则很少有修复的可能。修复方式可能主要依赖于切除修复。

5．交联　某些抗生素(如丝裂霉素 C)和某些试剂(如亚硝酸)能引起链内碱基交联和链间碱基交联，这就引起了双螺旋变形和阻碍复制时双链的分离。这种损伤也主要依赖于切除修复。

在 DNA 的各种待修复的缺陷中，值得注意的是紫外线所引起的胸腺嘧啶二聚体，其绝大部分是可以被修复的。细胞中，胸腺嘧啶二聚体引起细胞死亡的半致死数量为数千个，就是因为其绝大部分是可以被复制的，否则的话，只需要几个甚至只要一个二聚体就足以使细胞死亡。在实验中可以用酸水解 DNA 的方法分离胸腺嘧啶二聚体以测定修复状况。

上面讲述了好几类修复系统，似乎是能修复一切损伤。然而实际上当射线和化学物质作用时，细胞大批死亡。这又是什么原因呢？一个原因是 DNA 产生了不可修复的改变，如射线引起 DNA 双链在同一点上断裂；二是修复系统本身受到损伤；三是修复系统被饱和，修复不了大量的损伤；四是修复结果引起突变，因而不能产生有活性的对细胞生存所必需的产物。

以上关于 DNA 修复的各种机制的描述仍然是粗线条的，只涉及 DNA 修复的基本过程及主要的酶和蛋白质因子。这是因为目前研究的 DNA 损伤大多数是由射线以及烷化剂造成的。实际上体内许多生理和病理过程以及代谢产物都可能造成 DNA 的损伤。然而对于这类损伤及其修复机制还基本上没有进行详细的研究。其中值得注意的是真核生物蛋白质的多聚 ADP 核糖基化与 DNA 修复的关系。这个过程是由 ADP 核糖基转移酶(ADPRT)催化进行的。ADPRT 在 DNA 断裂作用的刺激下，迅速地从 $NAD^+$ 上将 ADP 核糖基转移到许多核蛋白上，其中包括 ADPRT 自己。经过核糖基化的蛋白质通过目前尚未知晓的途径促进DNA的修复。有人认为可能与 DNA 断裂部位的保护有关，有人认为可能与 DNA 连接酶 Ⅱ 的激活有关，并能解除组蛋白对于 DNA 连接酶 Ⅱ 的抑制作用(DNA 连接酶 Ⅱ 主要在修复过程中起作用)。然而以 N.Berger为首的小组对 ADP 核糖基化的研究结果表明，ADPRT 使大量的核蛋白产生 ADP 核糖基化，其中包括组蛋白，从而松解了染色质。由于ADP核糖基化作用很快耗尽了 $NAD^+$，则细胞中由该过程中释放出来的尼克酰胺与 PRPP 反应生成NMN，NMN 再与 ATP 反应生成 $NAD^+$。这样 PRPP 和 ATP 的贮备也急剧下降。在 DNA 复制一章中我们曾经讲述到 PRPP 和 ATP 是四种脱氧核苷三磷酸前体的生成所不可缺少的。因此细胞内四种前体分子浓度也急剧减少。N. Berger (1986)的实验表明，DNA 的损伤使 dGTP 含量迅速下降。这将严重地影响 DNA 的修复合成。因此 ADPRT 可能介导了 DNA 严重损伤情况下细胞的死亡。这与我们在 SOS 修复机制中所述的假设完全相反。在 SOS 修复过程中，大肠杆菌细胞以降低 DNA 合成的保真度为代价换取活命的机会；而这里细胞则宁死不屈。其差别就在于前者是原核生物，都是单细胞的，改变了 DNA 组成则能更好地适应环境，并通过 DNA 复制把这种改变固定下来；而后者是多细胞的真核生物，只有有机体作为整体存在时才有意义，个别细胞的 DNA 序列改变则有可能导致细胞恶性转化。ADPRT 的作用恰恰在于防止畸变的 DNA 修复结果被复制而固定下来。

## 第三节 限制与修饰[31~36]

### 一、限制-修饰现象

早在 50 年代初，Luria, Human, Bertani, Weigle 等人差不多同时发现了限制和修饰(restriction and modification)现象，当时他们称为寄主控制的专一性 (host-controlled specificity)。上述前二人是研究 T 偶数噬菌体的，而后二人是研究 λ 和 P2 噬菌体的。后来的研究表明，λ 噬菌体所表现出的寄主限制和修饰作用更具有代表性和普遍性。为了叙述的方便，我们先看一个实验数据(表 4-2))，然后就很容易明白限制作用和修饰作用的机理及其

相互关系。

表 4-2 *E.coli* λ 噬菌体的限制和修饰模式(表中数字代表相对种植效率)

| λ噬菌体 / *E.coli*菌株 | $\lambda_K$ | $\lambda_B$ | $\lambda_C$ |
|---|---|---|---|
| K | 1 | $10^{-4}$ | $10^{-4}$ |
| B | $10^{-4}$ | 1 | $10^{-4}$ |
| C | 1 | 1 | 1 |

从表 4-2 可以看出，从三种不同的 *E.coli* 菌株中繁殖出来的 λ 噬菌体的后代（分别用 $\lambda_K$、$\lambda_B$、$\lambda_C$ 表示)，再接种到同样菌株，接种效率都是 1 即 100%，这是不难理解的。然而如果接种到不同的菌株中，情况就不同了；$\lambda_K$ 在 B 株上接种效率只有 $10^{-4}$，$\lambda_B$在K株上也只 $10^{-4}$，$\lambda_C$ 在 K 株和 B 株上都为 $10^{-4}$，而不管 $\lambda_K$ 还是 $\lambda_B$ 在 C 株上的接种效率都为 1。表中所列仅是第一轮接种效率。如果将那些万分之一($10^{-4}$) 幸存者的后代噬菌体再接种到第一轮同样的*E.coli* 菌株中，如 $\lambda_K \rightarrow B \rightarrow B$，则接种效率也为 100%，然而，如果将万分之一幸存者的后代接种到祖辈噬菌体的寄主菌株中，如 $\lambda_K \rightarrow B \rightarrow K$，则接种效率也成为 $10^{-4}$，为什么？

原来，*E.coli* K 株和 B 株都各自含有一种限制修饰系统，其限制性内切酶能将外来的入侵者的 DNA 切断。每一种内切酶在 DNA 上都有一定的结合位点。当菌株本身的 DNA 也含有这样的结合位点时，则该菌株的甲基化酶就在这些识别位点上将腺嘌呤甲基化，成为 $N^6$-甲基腺嘌呤；或者将胞嘧啶甲基化，成为 5- 甲基胞嘧啶。这样，限制性内切酶就不会将自身的 DNA 降解掉。这两方面作用结合起来，就称为限制-修饰作用。当不同来源的噬菌体 λ 进入另一种大肠杆菌时，绝大部分被限制酶所降解，然而有极少数的 λ 噬菌体先被大肠杆菌的甲基化酶所修饰，因而不再受到大肠杆菌的限制酶的攻击而幸存下来，这些幸存者的后代因为有新寄主的修饰标记而缺少祖代寄主的修饰标记，因此只能感染新的寄主类型，而不再能感染祖代寄主。而大肠杆菌 C 株没有限制酶，因而其他来源的 λ 可以感染 C 株，而C株来源的 λ 不能感染 K 株和 B 株。这样，细菌借助于限制-修饰系统，就能区别自己 DNA 和外源 DNA。

这里需要指出，这里所谓自己 DNA 和外源 DNA 是对于它的限制-修饰模式而言，凡是限制-修饰模式相同的 DNA 均为自己DNA，这不仅包括同株系的不同个体的DNA，而且还包括寄生于这些菌株中的质粒和噬菌体。事实上，有些菌株的细胞染色体并没有限制-修饰系统的基因，而它的限制修饰系统是由其中的质粒或噬菌体的基因所编码的。也就是说，寄生的质粒或噬菌体赋于寄主以限制-修饰系统。因此有的不同菌株的差别不在于细菌本身，而在它所含的质粒或噬菌体的不同。在同一个细胞内的不同类型的 DNA，包括细胞DNA，质粒 DNA，噬菌体 DNA 等，都具有同样限制-修饰模式。这些同一细胞中的不同 DNA 可统称为该细胞中的居民 DNA(resident DNA)。

## 二、限制-修饰系统

据 Kessler 等人 1984 年统计，全世界已发现限制酶 600 多种，其中绝大部分为二类限制酶。大约三个菌株中就有一个含有 Ⅱ 类限制-修饰系统。根据限制-修饰系统的特性，可以分

为两大组，第一大组为二类限制酶(type Ⅱ)，它与相应的甲基化酶是分开的，不属同一个酶分子。在重组 DNA 技术和生化分析技术中所常用的限制性内切酶均为二类限制酶。第二大组包括一类限制酶和三类限制酶(typeⅠ、type Ⅲ)，这两类酶实际上应该称为Ⅰ类修饰-限制酶和Ⅲ类修饰-限制酶。它们的限制酶和甲基化酶都作为亚基包含在同一个酶分子中。除此之外，这三种不同的修饰限制酶还有许多其他不同的特征(见表 4-3)。

表 4-3 细菌中三种不同类型的修饰-限制酶

| | Ⅱ 类 酶 | Ⅰ 类 酶 | Ⅲ 类 酶 |
|---|---|---|---|
| 酶 分 子 | 内切酶和甲基化酶不在一起 | 三亚基双功能酶 | 二亚基双功能酶 |
| 识别位点 | 4～6bp 短序列大多数为回文结构 | 二分非对称序列 | 5～7bp 非对称序列 |
| 切割位点 | 在识别位点中或靠近识别位点 | 距离识别位点至少 1000bp，无特异性 | 在识别位点下游24～26bp处 |
| 限制反应与甲基化反应 | 分开的反应 | 互 斥 | 同时竞争 |
| 限制作用是否需要ATP | 不 需 要 | 需 要 | 需 要 |

Ⅱ 类酶的限制酶和甲基化酶是各自独立的两个酶，例如 EcoRI 的限制酶是相同亚基的二聚体，而其相应的甲基化酶为一单肽链。Ⅱ 类酶的识别位点一般为 4～6bp 的回文序列，这种对称性意味着这些位点的两条 DNA 链上均有可被甲基化的碱基。实际存在的甲基化状态有三种：全甲基化(两条链上都有甲基化碱基)，半甲基化(只有一条链上有甲基化碱基)，非甲基化(识别位点上没有甲基化碱基)。全甲基化的位点既不被限制；也不再被修饰。而半甲基化的位点也不被限制，但可以被修饰，成为全甲基化状态。非甲基化的位点在体外试验中既可作限制酶的底物，又可作甲基化酶的底物；但在体内的可能主要被限制(即被限制酶切断)。在居民 DNA 上，都是全甲基化的。随着 DNA 的复制，产生的子代 DNA 都是半甲基化的；但是不久，在复制过程中和复制结束后的短时间内就完成了全甲基化过程。而外源 DNA 的命运绝大多数是被限制，只有极少数能躲过限制作用而被修饰。二类限制酶的切割位点就在识别位点内或其附近。由于两条链上的切口位置不同。可以有粘性末端 (sticky ends 或 cohesive ends)和平齐末端(blunt ends 或flush ends)。粘性末端可以是 5′ 粘性末端，也可以是 3′ 粘性末端。有些 Ⅱ 类限制酶的识别位点的序列相同(切点不一定相同)，这些酶称为同位酶(isoschizomers)；有些Ⅱ类限制酶的作用后产生的 DNA 粘性末端相同(识别位点的序列不一定相同)，这些酶称为同尾酶(isocandamers)。大多数Ⅱ类限制酶具有严格的底物特异性，需要二倍旋转对称序列，即回文序列。但也有少数酶能作用于与回文序列相应的单链序列，或许这些单链序列之间能形成互补结构。或许是酶真的作用于单链，现在还不完全清楚。但是，这种作用比对双链识别位点的作用慢得多。拿研究较多的 Ⅱ 类限制酶 HaeⅢ 来说，两种作用速度相差 16 倍。也有人研究 Ⅱ 类限制酶对于 DNA-RNA 杂交分子的作用，有些酶能作用于杂交分子，在某些识别序列上(不考虑 d 和 r 差别)能切断 DNA，对 RNA 的作用是杂乱的，而且所需酶量是 DNA 双链的数十倍之多，在这种情况下很难排除酶制剂中混杂有 RNAase 的可能性。

表 4-4 列出常用的识别位点为 4 核苷酸对和 6 核苷酸对的 Ⅱ 类限制酶，其中有一部分限

制酶具有一种以上的识别位点。每一大方格的竖行(有五小格)中的几个酶互为同位酶，如 XmaI，SmaI 以及 AvaI。同一竖行中处于各大方格的相应位置上的酶则构成一组同尾酶，如 MboI，Bgl Ⅱ，BamHI，BclI。同尾酶产生相同的粘性末端，给 DNA 重组带来很大的方便。例如 Bgl Ⅱ 的识别位点为 AGATCT，BamHI 的识别位点为 GGATCC，但它们都产生相同的粘性末端 5′GATC，因而这两个酶所产生的 DNA 片断可以相互连接。但是它们产生新的序列，或者是 AGATCC，或者是 GGATCT。这两种新的序列既不能为 Bgl Ⅱ 所识别，也不能为 BamHI 所识别；但是另一个酶 BstYI 能识别这两种序列。

表 4-4 常用的 II 类限制酶的识别位点和切点

| | AATT | ACGT | AGCT | ATAT | CATG | CCGG | CGCG | CTAG | GATC | GCGC | GGCC | GTAC | TATA | TCGA | TGCA | TTAA |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| ▼□□□□ | | | | | | | | | Mbo I | | | | | | | |
| □▼□□□ | | Mae II | — | | | Hpa II | | Mae I | | HinP I | | | | Taq I | | Mse I |
| □□▼□□ | | | Alu I | | | | BstU I | | Dpn I | | Hae III | Rsa I | | | | |
| □□□▼□ | | | | | | | | | | Hha I | | | | | | |
| □□□□▼ | | | | | Nla III | | | | | | | | | | | |
| A▼□□□□T | | | Hind III | | Afl III | | Mlu I<br>Afl III | Spe I | Bgl II<br>BstY I | | | | | | | |
| A□▼□□□T | | | | | | | | | | | | | | Cla I | | Ase I |
| A□□▼□□T | | | | Ssp I | | | | | | Eco47 III | Stu I | Sca I | | | | |
| A□□□▼□T | | | | | | | | | | | | | | | | |
| A□□□□▼T | | | | | Nsp7524I | | | | | Hae II | | | | | Nsi I | |
| C▼□□□□G | | | | | Nco I<br>Sty I | Xma I<br>Ava I | | Avr II<br>Sty I | | | Eag I<br>Eae I<br>Gdi II | | | Xho I<br>Ava I | | Afl II |
| C□▼□□□G | | | | Nde I | | | | | | | | | | | | |
| C□□▼□□G | | | Pvu II<br>NspB II | | | Sma I | NspB II | | | | | | | | | |
| C□□□▼□G | | | | | | | Sac II | | Pvu I | | | | | | | |
| C□□□□▼G | | | | | | | | | | | | | | | Pst I | |
| G▼□□□□C | EcoR I | | | | | | BssH II | Nhe I | BamH I<br>BstY I | Ban I | | Asp718<br>Ban I | | Sal I | ApaL I | |
| G□▼□□□C | | Aha II | | | | | | | | Nar I<br>Aha II | | | Acc I | Acc I | | |
| G□□▼□□C | | | | EcoR V | | Nae I | | | | | | | Xca I | Hinc II | | Hpa I<br>Hinc II |
| G□□□▼□C | | | | | | | | | | | | | [illegible] | | | |
| G□□□□▼C | | Aat II | Sac I<br>Ban II<br>HgiA I<br>Bsp 1286 | | Sph I<br>Nsp7524 I | | | | | Bbe I<br>Hae II | Apa I<br>Ban II<br>Bsp1286 | Kpn I | | | Bsp1286<br>HgiA I | |
| T▼□□□□A | | | | | BspH I | BspM II | | Xba I | Bcl I | | Eae I | | | | | |
| T□▼□□□A | | | | | | | | | | | | | | BstB I | | |
| T□□▼□□A | | SnaB I | | | | | Nru I | | | Fsp I | Bal I | | | | | Dra I |
| T□□□▼□A | | | | | | | | | | | | | | | | |
| T□□□□▼A | | | | | | | | | | | | | | | | |

注意：表中少数酶有两个或两个以上的识别位点

I 类酶种类较少，可以用 EcoK 和 EcoB 为代表。它们的限制酶和甲基化酶各作为一个亚基存在于酶分子中，即 R 亚基和 M 亚基，另外还有一个负责酶分子对于 DNA 序列上特异靶位点识别的亚基，称作 S 亚基。R 亚基、M 亚基和 S 亚基分别由 *hsdR*、*hsdM*、*hsdS* 三个基因编码，它们属于同一个操纵元。EcoK 的组成是 $R_2M_2S$。对 EcoB 来说，它的 *hsd* 操纵

元有两个启动子，其顺序是 *p1*—*hsdR*—*p2*—*hsdM*—*hsdS*。两个启动子 *p1* 和 *p2* 是各自独立的，*p2* 启动子能使 M 基因和 S 基因独立于 R 基因之外单独表达。这大概就是 EcoB 酶分子组成不同于 EcoK 的原因。EcoB 的组成主要是 $R_2M_4S_2$，此外还有少量的不具限制酶活性的甲基化酶，其组成是 MS。当 MS 形式的甲基化酶在贮存期间可以出现 $M_3S$ 和 $M_2S$ 的形式，但是这可能并不代表细胞内状况。EcoK 和 EcoB 的亚基组成及分子量见表 4-5。

**表 4-5 I 类酶的物理性质**

| 酶 | 亚基分子量 | 可能的亚基组成 | S20, W | 酶分子量 |
|---|---|---|---|---|
| EcoK | R 135KDa<br>M 62KDa<br>S 52KDa | $R_2M_2S$ | 12 | 450KDa |
| EcoB | R 135KDa<br>M 60KDa<br>S 55KDa | $R_2M_4S_2$*<br>MS<br>$M_3S$<br>$M_2S$ | 11～18<br>6.6<br>11.3<br>8.0 | 450KDa～750KDa<br>105KDa<br>240KDa<br>175KDa |

* 表示细胞内的主要存在形式。

*hsdR* 基因突变体的表现型为 $R^-M^+$，即对 DNA 不能限制，只能修饰。*hsdS* 基因突变体的表现型为 $R^-M^-$，即既不能限制，也不能修饰。*hsdM* 基因的突变体的表现型亦为 $R^-M^-$，可见 M 亚基也参与 R 亚基的功能。M 亚基参与 R 亚基的功能，这是一种万一保安机制(fail-safe mechanism)，使 M 基因突变时不致于造成 $R^+M^-$ 这种致死的表现型。

大肠杆菌 EcoK 酶基因和 EcoB 酶基因互为等位基因。造成它们识别序列不同的原因在于 S 亚基的不同。一种菌株的 S 亚基可以和另一菌株的 R 亚基、M 亚基相结合而赋于它们新的位点识别能力。可见识别步骤是独立于后来的限制或修饰步骤。EcoB 和 EcoK 识别位点均由两段构成，一段为 3bp，另一段为 4bp，二者之间夹着一定距离的任意序列。EcoB 的识别序列是 $TGA(N)_8TGCT$，EcoK 的识别序列是 $AAC(N)_6GTGC$。EcoB 的甲基化位点已经清楚，由位于 $N_8$ 任意序列两侧的两个腺嘌呤(位于两条链上)，而 EcoK 的甲基化位点还不完全清楚，很可能处于与 EcoB 甲基化位点相似的位置，可以肯定的是其甲基化碱基亦为腺嘌呤。如果将 EcoK 识别序列两边各加一个 N(这两个任意碱基很可能是识别所必须的)，然后再与 EcoB 的识别序列对应起来，会发现它们七个不变的识别序列中有四个相同(粗实线处)：

```
EcoB, 5'TGA* NNNNNNNNTGCT
         ACT  NNNNNNNNA*CGA 5'

EcoK, 5'NAA° CNNNNNNGTGCN
         NTT  CNNNNNNCA°CGN 5'
```

其中 A* 为 EcoB 的甲基化位点，A° 为 EcoK 的可能的甲基化位点。从它们识别序列长度来看，酶分子只结合于 DNA 双螺旋的一侧。

I 类酶与其识别位点结合以后，便根据该位点的甲基化状况决定酶分子行使哪种功能，或者限制，或者修饰，或者既不限制又不修饰(图 4-15)。无论行使哪种功能，酶分子中的 M 亚基必须首先和辅因子 S-腺苷甲硫氨酸(SAM)结合，使酶分子变构而处于活化状态，以便与识别位点结合。这就是 M 亚基参与限制功能的机制。M 亚基和 S 亚基的突变以及 SAM 的缺乏

均影响到酶分子的变构作用，使表现型均为 $R^-M^-$。这就是上面所说的万一保安机制。酶-SAM 与 DNA 结合后接着与 ATP 反应。如果结合位点是完全甲基化的，则 ATP 的存在使酶-SAM 脱离 DNA；如果结合位点是半甲基化的，则非甲基化的链可被酶-SAM 甲基化，同时 SAM 转变为 S- 腺苷高半胱氨酸 (SAH，S-adenosyl-homocysteine)。甲基化作用可以不需要 ATP，但在 ATP 存在时反应效率提高。如果结合位点是非甲基化的，ATP 可以使酶分子再次变构而表现出限制酶活性，同时释放出 SAM。在缺少ATP时，酶可将非甲基化的序列甲基化，这在体外用 EcoK 已得到证明。这一次变构作用与 R 亚基有关。限制酶在切割 DNA 时，需要水解 ATP。切割后，酶分子脱离 DNA，回到原来的非活跃形式。

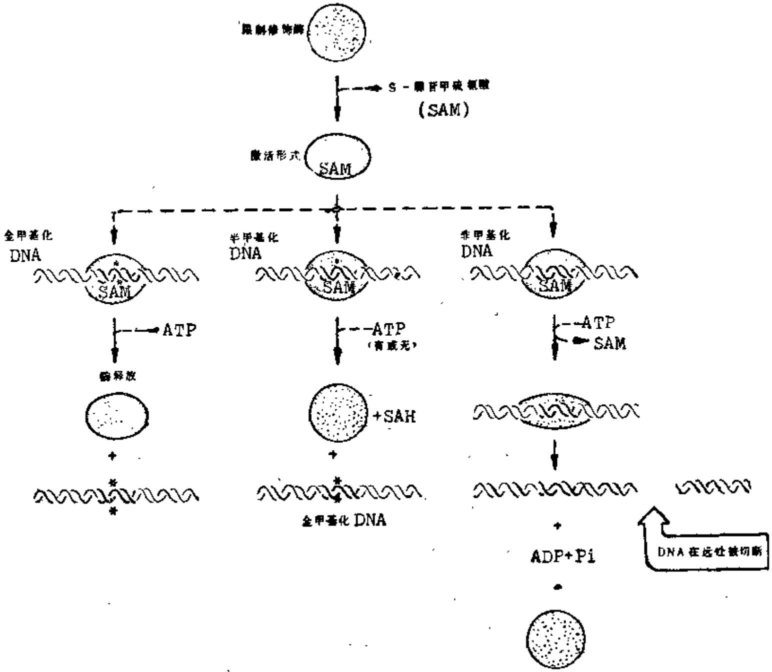



图 4-15 I 类酶的限制-修饰机制

I 类限制酶的切割位点相距 1 000 bp 以上，切割位点的序列特异性不严格，但也不是完全随机的。有些序列可以优先被切割。因此要鉴定 I 类酶的识别序列就不能采用分析切口处序列的办法，而只能通过分析甲基化序列以及分析妨碍识别的 DNA 突变位点。I 类酶是如何识别其切割位点？目前有两种假说。一种是酶移动假说，即酶分子在识别位点结合以后，沿着 DNA 移动，由于某种机制在切割位点上停下来。这种假说的证据较少。而第二种假说是 DNA 移动假说，即酶分子有两个结合位点，分别与识别序列上的两部分特异性的碱基相结合(如 EcoB 的 TGA……TGCT)而不与中间的任意序列结合。然后在 ATP 存在时，酶分子再次发生变构，这时酶分子只与一部分识别序列相结合(可能是四个碱基的序列)，而酶分子用某

种尚不知晓的机制将另一边的 DNA 双螺旋向中间拉过来，通过分子上不再结合 DNA 的一个结合位点的检查，一直到发现其切割位点为止，将DNA的一条链切断。然后大量消耗ATP，释放出总共 70b 左右的寡聚核苷酸片断。然后可能由第二个酶分子将 DNA 的另一条链切断。其过程见图 4-16。电子显微镜已观察到与酶结合的DNA啫噗，是对DNA移动假说的支持。

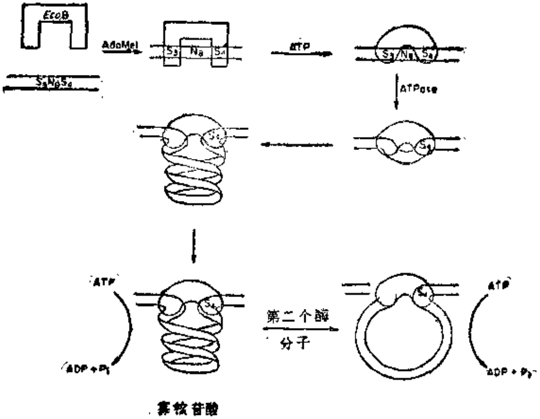



图 4-16　I 类限制酶切割机制的 DNA 移动假说模型

Ⅲ 类酶发现于 70 年代末。大肠杆菌质粒 P1 和 P15 编码的 EcoP1 和 Ecop15 以及血清型流感嗜血杆菌 Rf 中的 Hinf 均为 Ⅲ 类酶。Ⅲ 类酶由 R 亚基和 MS 亚基这两种亚基所组成。R 亚基负责限制反应，而 MS 亚基负责识别和修饰反应。R 亚基基因突变，可产生 $R^-M^+$ 表现型。而 MS 基因的突变，如发生在负责识别的区域，则产生 $R^-M^-$ 表现型；如果发生在负责修饰的部分，则会产生 $R^+M^-$ 而引起致死。通过比较 P1 和 P15 系统的 *hsdMS* 基因结构，负责识别功能位于基因的中部，R 亚基分子量 106 000～110 000 道尔顿，MS 亚基为 73 000～80 000 道尔顿。

Ⅲ类酶与识别位点的结合过程是严格需要 ATP 的，但又无可测得的 ATP 水解。在与识别位点结合之后，出现修饰作用和限制作用的竞争。修饰作用在识别位点上进行，这与 MS 亚基同时具有修饰和识别功能是一致的。限制切点出现在识别位点的一侧 24～26bp，可能是因为该酶有足够的大小能同时与识别位点和切割位点结合。

Ⅲ类酶和 I 类酶一样，也是在腺嘌呤上甲基化，其甲基供体亦为 SAM。然而，EcoP1 和 EcoP15 的识别位点序列中都只有一条链上有 A 存在。EcoP1 的识别位点为 AGACC，EcoP15 为 CAGCAG。既然如此，甲基化是如何在 DNA 复制过程中保持下去的呢？由于 DNA的半保留复制，一个子代 DNA 分子与亲代一样，在一条链上甲基化；而另一个子代DNA 则完全没有甲基化，因而可能成为限制和修饰的对象。什么因素决定这个序列被限制还是被修饰，现在还不清楚。但人们设想，其修饰作用是与复制作用紧密相联的，当细胞 DNA 进行复制时，条件有利于修饰作用的进行，而外来 DNA 的主要命运是被限制降解。

### 三、限制修饰系统的生物学意义

本节一开头我们就讲解了细菌为对付外来的 DNA 和保护自己细胞内的 DNA，演化出限制-修饰酶系。最早的一批例子就是噬菌体 λ 和 P2 的感染和转导受到了限制，后来相继发现 *E.coli* 的 Hfr 菌株和 $F^-$ 菌株之间的接合也受到阻碍，还有原生质球（spheroplasts）对于 DNA 的摄取也受到寄主限制-修饰特异性的控制。

然而，限制-修饰系统如果只有此功能的话，I类酶的组成和作用机制似乎没有必要那么复杂，切点也没有必要一定在1 000 bp开外。人们的印象是，切割下来的 DNA 似乎对细胞还有用处。I 类酶很可能还具有一种目前尚不知晓的功能。如非特异性重组等。Ⅱ类酶、Ⅲ 类酶常常是由质粒或溶原性噬菌体所编码，Roberts 已指出，它们的功能除了限制作用以外，还可能促进位点特异性重组，促进质粒之间、原噬菌体之间的遗传信息的交流(shuffling)。此外，限制酶还可能与其他核酸酶配合，为细胞合成 DNA 提供核苷酸前体。

## 提　　要

DNA 是遗传信息的载体，它需要有极高的保真度，这不仅有赖于复制体系的完善，而且还需要有能够纠正业已存在的错误这样一类修复体系。尿嘧啶糖基酶系统和错配修复系统与纠正复制错误直接相关，因此称为复制修复；但其他因素所产生的这类错误可同样被修复。

生物体内外许多因素都能造成 DNA 分子结构的损伤，其中以胸腺嘧啶二聚体的形成和修复机制研究得最为清楚。光复活，切除修复，重组修复，SOS修复以及二聚体糖基酶修复系统为胸腺嘧啶二聚体的五种主要修复系统。其中光复活是唯一利用光能的修复系统，而其余均以 ATP 为能源；光复活，切除修复和二聚体糖基酶修复都是修复模板链，SOS 修复是产生连续的子链。SOS 修复是唯一导致突变的修复。

对于其他类型的 DNA 损伤，也存在许多修复系统。非标准碱基可由相应的糖基酶切除。糖基酶作用后的 AP 位点和碱基丢失产生的 AP 位点都可由 AP 内切酶等进行修复。嘌呤碱基插入酶可在 AP 位点上插入相应的嘌呤碱基。烷基化损伤则由一类更为复杂的修复系统进行修复。

细菌为了防止外源 DNA 的入侵以及对外源 DNA 的利用，演化出限制酶，而同时又对自己 DNA 上的相应限制酶位点进行修饰(甲基化)，从而使自己的 DNA 不被自己的限制酶所破坏。微生物中有三类修饰-限制酶，其中Ⅱ 类酶的内切酶和甲基化酶不在同一个分子中，限制作用不需要 ATP。Ⅱ 类酶广泛用于 DNA 重组技术。人们对于限制-修饰系统的生物学意义还没有完全明白，可能在重组过程中也有重要作用。

## 参　考　文　献


1. Sancar, A. and Sancar, G.B. (1988) "DNA Repair Enzymes." *Ann. Rev. Biochem.* 57, 29～67.
2. Modrich, P., (1987) "DNA Mismatch Correction." *Ann. Rev. Biochem.* 56, 435～466.
3. Wagner, R. and Meselson, M. (1976) "Repair Tracts in Mismatched DNA Heteroduplexes." *Proc. Nat. Acad. Sci.* 73: 4135～4139.
4. Lu, A.-L. et al. (1983) "Methyl-directed Repair of DNA Base-pair Mismatches in Vitro" *Proc. Nat. Acad. Sci.* 80-4639～4643.
5. [illegible], M. et al. (1987) "Mismatch Repair and Recombination in *E.coli*." *Cell* 50: 621～626.

6. Singer, B. and Kusmierek, J.T. (1982) "Chemical Mutagenesis." *Ann. Rev. Biochem.* 52: 655〜693.
7. Haseltine, W. (1983) "Ultraviolet Light Repair and Mutagenesis Revisited." *Cell* 33: 13〜17.
8. Friedberg, E. C. et al. (1981) "Enzymes that Incise Damaged DNA." in *The Enzymes* (ed. Boyer, P.D.) 14, 251〜279.
9. Sutherland, B. M. (1981) "Photoreactivating Enzymes." in *The Enzymes* (ed. Boyer, P. D.) 14: 481〜515.
10. Sancar, G. B. and Rupp, W. D. (1983) "A novel Repair Enzyme: UvrABC Excision Nuclease of *Escherichia coli* Cuts a DNA Strand on Both Side of the Damaged Region." *Cell* 33: 249〜260.
11. Yeung, A.T. et al. (1983) "Enzymatic Properties of Purified *Escherichia coli* UvrABC Proteins." *Proc. Nat. Acad. Sci.* 80: 6157〜6161.
12. Gruskin, E. A. and Lloyd, R.S. (1988) "Molecular Analysis of Plasmid DNA Repair Within Ultravioletirradiated *Escherichia coli*. I. T4 Endonuclease V-initiated Excision Repair." *J. Biol. Chem.* 263, 12728〜12737.
13. Gruskin, E.A. and Lloyd, R.S. (1988) "Molecular Analysis of Plasmid DNA Repair within Ultravioletirradiated *Escherichia coli*. II. UvrABC-initiated Excision Repair and Photolyase-catalyzed Dimer Monomerization." *J. Biol. Chem.* 263, 12738〜12764.
14. Hannan, M. A. and Nasiam, A. (1984) "Repair mechanisms and Mutagenesis in Microorganisms" in *Repairable Lesions In Microorganisms*, ed. Hurst, A. and Nasim, A., Academic Press, pp. 125〜148.
15. Radding, C. (1982) "Homologous Pairing and Strand Exchange in Genetic Recombination." *Ann. Rev. Genetics* 16: 405〜437.
16. Szostak, J. et al. (1983) "The Double-strand Break Repair Model for Recombination." *Cell* 33: 25〜35.
17. Walker, G. (1984) "Mutagenesis and Inducible Responses to Deoxyribonucleic Acid Damage in *Escherichia coli*." *Microbiol. Revs.* 48: 60〜93.
18. Little, J. et al. (1980) "Cleavage of the *Escherichia coli* Lex A Protein by the RecA Protease." *Proc. Nat. Acad. Sci.* 77, 3225〜3229.
19. Sancar, G. et al. (1982) "The *uvrB* Gene of *Escherichia coli* Has Both LexA-repressed and LexA-independent Promoters." *Cell* 28 523〜530.
20. Kato, T. and Shinoura, Y. (1977) "Isolation and Characterization of Mutants of *Escherichia coli* Deficient in Induction of Mutations by Ultraviolet Light." *Mol. Gen.* Genetics 156, 121〜131.
21. Elledge, S. and Walker, G. (1983) "Proteins Required for Ultravolet Light and Chemical Mutagenesis, Identification of the Products of the *umuC* Locus of *Escherichia coli*" *J. Mol. Biol.* 164, 175〜192.
22. Duncan, B.K. (1981) "DNA Glycosylase" in *The Enzymes* (ed. Boyer, P.D.) 14, 565〜586.
23. Livneh, Z. and Sperling, J. (1981) "DNA Base-insertion Enzymes (Insertases)" in *The Enzymes* (ed. Boyer, P. D.) 14, 549〜563.
24. Samson, L. and Cairns, J. (1977) "A New Pathway for DNA Repair in *Escherichia coli*." *Nature* 267: 281〜282.
25. Lindahl, T. et al. (1988) "Regulation and Expression of the Adaptive Response to Alkylating Agents." *Ann. Rev. Biochem.* 57: 133〜157.
26. Lindahl, T. et al. (1982) "Suicide Inactivation of the *E. coli* $O^6$-methylguanine-DNA Methyltransferase." *EMBO J.* 1: 1359〜1363.
27. Teo, I. et al. (1984) "Induction of Resistance to Alkylating Agents in *E. coli*; The *ada* Gene Product Serves both As a Regulatory Protein and As an Enzyme for Repair of Mutagenic Damage." *EMBO J.* 3: 2151〜2157.
28. Teo, I. et al. (1986) "The Intracellular Signal for Induction of Resistance to Alkylating Agents in *E. coli*." *Cell* 45, 315〜324.
29. van Duin, M. et al. (1986) "Molecular Characterization of the Human Excision Repair Gene ERCC-1: cDNA Cloning and Amino Acid Homology with the Yeast DNA Repair Gene RAD 10." *Cell* 44: 913〜923.
30. Gaal, J.C. et al (1987) "Cellular Euthanasia Mediated by a Nuclear Enzyme; a Central Role for Nuclear ADP-ribosylation in Cellular Metabolism." *TIBS* 12: 129〜130.
31. Arber, W. and Linn, S. (1969) "DNA Modification and Restriction." *Ann. Rev. Biochem.* 38: 467〜500.
32. Smith, J. D. et al. (1972) "Host Specificity of DNA Produced by *E. coli*. XIII. The Role of Nucleotide Methylation in *in Vivo* B-specific Modification." *J. Mol. Biol.* 63: 1〜8.
33. Yuan, R. and Hamilton, D. L. (1984) "Type I and Type III Restriction-modification Enzymes." in *DNA Methylation*, ed. Razin, A. et al. Springer-Verlag, New York, pp. 11〜37.
34. Smith, H. O. and Kelly, S.V. (1984) "Methylases of the Type II Restrion-modificatiction Systems." in *DNA Methylation*, ed. Razin, A. et al., Springer-Verlag, New York, pp. 39〜71.
35. Yuan, R. and Smith, H. O. (1984) "The Ristriction and Modification DNA Methylases, An Overview." in *DNA Methylation*, ed. Razin. A. et al., Springer-Verlag, New York, pp. 73〜80.
36. Grafstrom, R. H. et al. (1984) "DNA Methylation, DNA Replication and Repair." in *DNA Methylation*, ed. Razin, A. et al., Springer-Verlag, New York, pp. 111〜126.

# 第五章 突　　变

## 第一节 概述：突变定义及其分类[1,2,3]

什么叫突变(mutation)？突变是一种遗传状态。可以通过复制而遗传的 DNA 结构的任何永久性改变，都叫突变。所有的突变都是 DNA 结构中碱基所发生的改变。携带突变的生物个体或群体或株系，叫做突变体(mutant)。正是由于变突体中 DNA 碱基序列的改变，因而产生了突变体的表现型。突变位点可能存在于基因内，该基因称为突变基因(mutant gene)。没有发生突变的基因称为野生型(wild type)基因。例如，负责细菌合成 Arg 的基因 *arg* 为野生型基因。如果该基因突变而失去了合成 Arg的能力，就必须由外界供应 Arg，否则细菌就不能生长。细菌的这种表现型就称为 Arg⁻ 表现型，即精氨酸合成缺陷表现型，其基因型写作 $arg^-$。由此可见，突变体是指有机体的表型特征中有一种(或多种)与野生型个体的该特征有所不同。具有这样遗传状态的有机体就叫做突变体。需要指出的是，所谓野生型是指有机体的正性状，如能够分解某种底物的能力，能够合成某种物质(如氨基酸)的能力。在多数情况下，从自然界分离得到的有机体种类都具有这种正性状。但有时并不是这样，例如大家最熟悉的 *E·coli lac* 基因，通常从自然界分离到 *E·coli* 都是 $lac^-$，即不能利用乳糖的种类。然而，我们仍然把 $lac^+$ 称为野生型，而把 $lac^-$ 称为突变体。从这一点上来说，"野生型"这一名词常常是误用名称，但它已为遗传学家所习用。现在，我们只要遵循上述定义，也就不会再望文生义了。关于野生型和突变体的表现型和基因型的表示方法列于表 5-1。所有的基因型名称均用 3 个小写的斜体字母或小写字母在底下划线，而有关的具体基因则在 3 个小写字母后用大写的斜体字母表示，如 $lacZ^+$、$lacZ^-$。所有的表现型名称均用 3 个正写的字母表示(其中第

**表 5-1　细菌突变体和突变的表示方法**

| 表现型或基因型 | 符　　号 |
|---|---|
| 表现型： | |
| 缺乏合成或利用某种物质的能力 | $Sub^-$ |
| 具有合成或利用某种物质的能力 | $Sub^+$ |
| 对某种抗菌素的抗性 | Ant-r |
| 对某种抗菌素的敏感性 | Ant-s |
| 基因型： | |
| 能够合成或利用某种物质的野生型基因 | $sub^+$ |
| 影响合成或利用某种物质的突变基因 | $sub^-$ |
| 突变的 *subA* 基因 | $subA^-$ |
| 具有温度敏感表现型 *subA* 基因的突变 | *subA*--(*Ts* |
| 具有琥珀表现型的 *subA* 基因的突变 | *subA*-(*Am*) |
| *subA* 基因的 63 号突变 | *subA*63 |
| 对于某种抗菌素的抗性或敏感性的基因 | *ant* |
| 对于某种抗菌素抗性的基因型 | *ant-r* |
| 对于某种抗菌素敏感性的基因型 | *ant-s* |

一个字母大写)，如 $Lac^+$，$Lac^-$。如果是对抗菌素的抗性(resistance)或敏感性(sensitivity)则在后面加上 -r 或 -s。如果阿拉伯数字，则表示突变体分离的时间顺序。如果是温度敏感突变，则在基因型符号后加写(*Ts*)，如果是某种无义突变，则在基因型符号后加写该无义突变的符号：(*Am*)、(*Oc*)或(*Opal*)，*Am*为琥珀型(amber)，密码子为 UAG；*Oc* 为赭石型(Ochre)，密码子为 UAA；*Opal* 为乳石型，密码子为 UGA。

引起突变的物理因素(如X-射线)和化学因素(如亚硝酸盐)，称为突变剂(mutagen)。由于突变剂的作用而产生突变的过程或作用称为突变生成作用(mutagenesis)，中文有时亦简称突变。如果这一作用是在自然界中发生的，不管是由于自然界中突变剂的作用的结果还是由于偶然的复制错误被保留下来，都叫做自发突变生成(spontaneous mutagenesis)。其结果是产生一种称为自发突变(spontaneous mutation)的遗传状态，携带这种遗传状态的个体或群体或株系就称为自发突变体(spontaneous mutant)。自发突变的频率平均为每一核苷酸每一世代 $10^{-9}$～$10^{-10}$。相反，如果这种作用是由于人们使用突变剂处理生物体而产生的，就叫做诱发突变生成，简称诱变(induced mutagenesis)。它产生的遗传状态叫做诱发突变(induced mutation)，携带这种遗传状态的有机体就叫做诱发突变体(induced mutant)。诱发突变的频率高得多(千倍以上)，而且现在可以在离体条件下进行定向诱变。

以上我们对突变及其有关的概念作了定义。下面，我们将把突变从不同的角度进行分类。实际上，突变是无法进行单系统分类的，因为它的因、果、状态、过程诸方面是既有区别又有联系的。与其说是分类，还不如说是继续进行定义和解释，就像上面所说的，按照突变生成的过程可以分为自发突变和诱发突变，如此之类。

从 DNA 碱基序列改变多少来分，可以分为单点突变(point mutation)，即只有一个碱基对发生改变，和多点突变(multiple mutation)，即有两个或两个以上的碱基对发生改变。点突变可以是碱基替代 (base substitution)，碱基插入(base insertion)或碱基缺失(base deletion)。但点突变这个术语常常是指碱基替代。碱基替代可以分为两类，一类叫转换(transition)，即嘌呤到嘌呤或嘧啶到嘧啶的变化；另一类叫颠换(transversion)，即嘌呤到嘧啶或嘧啶到嘌呤的变化。点突变的重要特点之一是它具有很高的回复突变率。缺失突变和插入突变通常指较长的碱基序列的缺少或增加。这样的突变，特别是缺失突变，其回复突变率极低。有时，插入序列本身携带遗传信息，如转座成分。转座成分的插入可使插入位的基因失活，同时又带进新的基因。插入序列可以切离失去，准确的切离可使突变基因又回复成野生型基因(详见 DNA 重组一章)。

从对阅读框架的影响来看，有所谓移框突变(frameshift mutation)。插入、缺失一个或两个碱基都能引起移框突变；扁平的碱性染料分子的嵌合也常引起移框突变。移框突变不但改变了产物的氨基酸组成，而且会出现蛋白质合成的过早的终止。如果移框突变发生在必需基因上，常常是致死的。如果插入或缺失三个碱基，则阅读框架不变，其产物常常有活性或有部分活性。

从对遗传信息的改变来看，点突变中的碱基替代突变可以进一步分为同义突变(synonymous mutation)、错义突变(missense mutation)和无义突变(nonsense mutation)。同义突变是指没有改变产物氨基酸序列的密码子变化，显然这是与密码子的简并性相关的。错义突变是指碱基序列的改变引起了产物氨基酸序列的改变。有些错义突变严重影响到蛋白质活性甚至完全无活性，从而影响了表现型。如果该基因是必需基因，则该突变为致死突变

(lethal mutation)。也有不少错义突变的产物仍然有部分活性,使表现型介于完全的突变型和野生型之间的某种中间类型。这样的突变又称为渗漏突变(leaky mutation)。有一些错义突变不影响或基本上不影响蛋白质活性,不表现出明显的性状变化,这种突变常被称为中性突变(neutral mutation)。中性突变连同同义突变一起常被称为无声突变(silent mutation)。无义突变是指某个碱基的改变使代表某种氨基酸的密码子变为蛋白质合成的终止密码子,如赖氨酸的密码子 AAG 突变为终止密码子 TAG(即 UAG)。无义突变还可以由其他各种突变产生,如移框突变、插入突变和缺失突变都能造成无义突变。无义突变使肽链过早终止,因而蛋白质产物一般是没有活性的。但是由点突变中的碱基替代突变产生的无义突变,如果发生在靠近 3′ 末端处,它所产生的多肽链常有一定的活性,表现为渗漏型。这类多肽还有一个特点,它们多半已具有野生型多肽链的抗原特异性。这样,用野生型基因产物的抗体作免疫学反应就可以检定这些不完全多肽链的存在,这种方法在无表型性状可利用时显得更为重要。

从突变表现型对外界环境的敏感性来区分,可分为非条件型突变(nonconditional mutation)和条件型突变(conditional mutation)。最常见的条件型突变为温度敏感突变(temperature-sensitive mutation)。第三章中所说的 *dna* 突变体均为温度敏感突变体。

如果从突变的效应背离或返回到野生型这两种方向上来讲,可以分为正向突变(forward mutation)和回复突变(back mutation 或 reverse mutation)。正向突变是指改变了野生型性状的突变。突变体所失去的野生型性状可以通过第二次突变得到恢复,这种第二次突变就叫做回复突变。真正的原位回复突变很少(也就是回复到野生型的 DNA 序列),而大多数是第二点突变。原来的突变位点依然存在,而它的表现型效应被基因组第二位点的突变所抑制,因而又称为抑制突变(suppressor mutation)。抑制突变的机制将在第三节中叙述。

突变位点也可能存在于负责基因调控的 DNA 序列当中。比如存在于启动子区域。这类突变中,有的能增强启动子对于转录的发动作用,就称为启动子上升突变(promoter up mutation);有的突变则降低启动子的效能,称为启动子下降突变(promoter down mutatoin)。如果突变位点发生在操作子(operator)上,其位点不能为阻遏蛋白所识别;或者由于调节基因发生突变,不能产生功能的阻遏蛋白。这两种情况或二者之一都使结构基因失去了负向控制,产生不依赖于需要,在细胞中有固定数量的蛋白质。基因的这种表达方式叫做组成型表达;产生这种表达方式的操作子突变或调节基因的突变就叫做组成型突变(constitutive mutation)。启动子突变体和组成型突变体是研究基因调控的重要手段。

## 第二节 条件型突变

基因组中的任何基因都可以出现条件型突变。对于非必需基因,我们可以选择条件突变也可以选择非条件突变来研究它们的功能。而对于任何情况下都是必需的基因,任何使之失去活性的突变都必然引起细胞死亡,因而只能通过选择条件突变体来研究它们。条件突变体表现为条件致死,即在许可条件(permissive condition)下细胞生长正常或近乎正常,而在非许可条件(nonpermissive condition)下表现出病态或死亡。生物体任何基因的表达都是依赖于各种各样的体内条件和环境条件的,所以从广义来说,任何突变体都可以看作是条件型的。然而在实际的研究工作中,人们只选择比较容易控制的条件,如温度、必需的营养物质、抗菌素等。对温度敏感突变体的研究最多,因为通过提高或降低温度,可以很快很方便地关闭或

开启某一基因产物的活性。温度敏感突变体已被用来研究基因尤其是必需基因在细胞生存、生长、分化和个体发育中的作用。一般来说，细菌温度敏感突变体的许可温度为30°左右，非许可温度为42℃左右，也有少数突变体的非许可温度低于许可温度。温度敏感突变一般为错义突变。改变了的蛋白质产物在许可温度下活性正常或接近于正常，而在非许可温度下无活性。这可能是由于氨基酸的改变在非许可温度下不能形成高级结构或多亚基复合物。

## 第三节　回复突变和抑制突变[4~17]

### 一、回复突变的鉴定和分类

前面我们已经说过，真正的原位回复突变很少，大多数回复突变都是第二点突变抑制了第一次突变造成表现型(表型抑制)，使得野生表现型得以恢复或部分恢复。真正的回复突变正好发生在原来位点，使突变基因回复到与野生型基因完全相同的 DNA 序列。比如说，原来野生型基因的一段序列为 ACTG；在出现 C→T 转换后，该序列成为 ATTG；则该基因产物没有活性，如果在以后的某一世代中第一个 T 出现 T→C 的转换，则该序列又变为 ACTG，野生表现型也就随之恢复。回复突变可以自发地产生，其频率一般是正向突变的十分之一左右，也可以用突变剂处理增加其频率。并非所有的正向突变都可以自发地回复到野生型状态。对于双重突变体(含有两个决定同一性状的突变位点，这两个位点可以在一个基因上或在两个作用相关的基因上)，其回复突变发生率非常之低(两个单点回复突变的乘积)，一般低于 $10^{-12}$，在实验中往往测定不出。另外，对于大片断的缺失突变，要在原位插入同样长度和序列的碱基，因而使产物的氨基酸序列恢复到野生型结构，或者插入碱基的长度和序列虽与失去的不同，但其产物具有野生型产物功能，这样的几率几乎为 0，也就是说基本上不可能有回复突变。反过来，如果一个突变在使用各种诱变剂反复处理后仍然无回复突变发生，则很可能是双重突变体，多重突变体或缺失突变体。由于大多数回复突变都不是真正的原位回复突变，因此鉴定回复突变主要不是根据其基因型而是依据其表现型。

第二点回复突变并没有改变正向突变的 DNA 碱基序列，只是其突变效应被抑制了。因而第二点回复突变通常都称为抑制突变(suppressor mutation)。抑制突变可以发生在正向突变的基因之中，叫做基因内抑制突变(intragenic suppressors)；也可以发生在其它基因之中，这叫做基因间抑制突变(intergenic suppressors)。根据野生表型恢复作用的性质还可以分为直接抑制突变(direct suppressors)和间接抑制突变 (indirect suppressors)。直接抑制突变是通过恢复或部分恢复原来突变基因蛋白质产物的功能而使表现型恢复为野生型状态。所有基因内抑制突变的作用都是直接的。一些改变翻译性质的基因间抑制突变的作用也是直接的。间接抑制突变不恢复正向突变基因的蛋白质产物的功能，而是通过改变其他蛋白质的性质或表达水平而补偿原来突变造成的缺陷，从而使野生表型得以恢复。

### 二、基因内抑制突变

由碱基替代引起的错义突变和由插入或缺失一个或几个碱基引起的移框突变都可以被基因内另一位点的突变所抑制。由于处在同一基因中，正向突变和抑制突变一般紧密连锁，距离不是很大，在细菌中至多数千 bp，在真核生物中由于内元的存在，可能高达数万 bp，但大多数相距不远，甚紧密相邻。在不是很精细的遗传学分析中很难测定出二者之间的重组，于是常

把基因内抑制突变当成真正的原点回复突变。对于重组频率较低而又较难得到大量杂交后代的真核生物，几乎不可能通过常规的遗传学方法辨别真正的原位回复突变和相距几个 bp 的基因内抑制突变。必须采用 DNA 序列分析或蛋白质序列分析才能鉴别出来。

错义突变所造成的野生表型的丧失，部分原因可能是影响到蛋白质空间结构的形成，也就是影响到突变点氨基酸残基与分子内其他某一部位的氨基酸残基的相互作用。这种作用可能是正负电荷之间的静电吸引作用，也可能是疏水作用或者是生成氢键。若是静电作用，只要恢复到两点之间电荷相反即可(图 5-1a)。而图 5-1b 表示由六个氨基酸残基形成的疏水作用。由于正向突变使较小的氨基酸残基(如丙氨酸)变为较大的氨基酸残基(如苯丙氨酸)，而回复突变使原来与丙氨基酸残基作用的较大氨基酸残基变为较小的氨基酸残基，恢复了疏水作用力，因而恢复了蛋白质分子的功能。

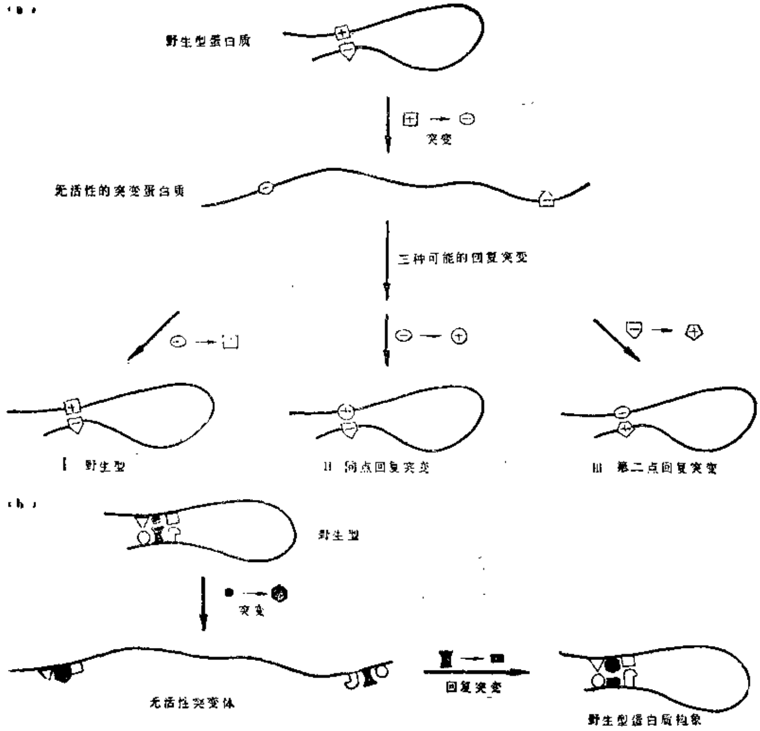



图 5-1 错义突变的基因内抑制突变

对于第二点氨基酸取代抑制突变体的分离和研究在决定蛋白质的三维结构中非常重要。通常，如果 A 氨基酸被 X 氨基酸取代而引起蛋白质失活可以被取代 B 氨基酸的 Y 氨基酸所补偿，则 A、B 两个氨基酸在三维结构中要么相邻，要么处在两个彼此作用的区域中。

有些基因内抑制突变表现为温度敏感型，这是不难理解的，因为第二点的氨基酸取代虽然

可以在许可温度下与正向突变的氨基酸互相作用，但这种作用没有野生型那对氨基酸之间相互作用的适宜和稳定，在较高温度下易破坏。

在图 5-1aI、II 中，I 为真正的原点回复突变，即恢复到野生型的碱基序列。II 的突变位点与正向突变影响同一位置的氨基酸，如果影响的碱基位置不同，可以算作基因内抑制突变；如果改变的是同一个碱基，不管野生型是否恢复，都不应称为抑制突变，因为此时正向突变已经不存在了。

移框突变的回复突变几乎全是基因内第二点突变，即基因内抑制突变。不难理解，任何碱基替代对移框突变都毫无用处，只能以错改错，即由新的移框突变来校正(图 5-2)。显然，由于两个突变点之间的氨基酸均与野生型不同，因而两点之间的距离越小，对蛋白质结构和功能的影响就越小，对正向突变的表现型的抑制作用也越大。否则，即使阅读框恢复，但由于蛋白质序列改变太多，野生性状就不能恢复。其次，即使两个突变点距离很近，两点之间的那几个氨基酸也要经得起变化，否则一个氨基酸的变化有时就足以使蛋白质失活。

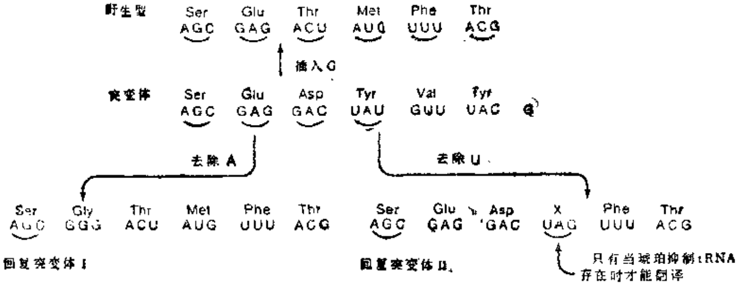



图 5-2 移框突变的抑制突变

突变剂可以提高基因内抑制突变的频率。对于碱基替代突变，应该采用产生碱基替代突变的突变剂，如 5-溴尿嘧啶 (BU)，2-氨基嘌呤(AP)；对于移框突变，则应采用产生移框突变的突变剂，如吖啶类染料。如果上述两类不同的正向突变与两类不同突变剂交叉配合，则不能产生回复突变。这是遗传学中鉴定碱基替代突变还是移框突变的重要标准。在碱基替代突变中，如果正向突变是由 5-溴尿嘧啶或 2-氨基嘌呤等引起的由 A·T 到 G·C 的突变，则各突变剂的回复突变作用顺序是：羟胺 >BU；若正向突变是由羟胺等引起的G·C 到 A·T的突变，则回复突变作用顺序是 BU>AP> 羟胺。

### 三、基因间抑制突变

正向突变的无义突变，错义突变和移框突变都可为另一基因上的突变所抑制，这些抑制突变分别称为无义抑制突变(nonsense suppressor mutation)，错义抑制突变(missense suppressor mutation)，移框抑制突变(frameshift suppressor mutation)。这类抑制突变均发生在 tRNA 基因或与 tRNA 功能有关的基因上。这些基因又称为抑制基因(suppressor genes)。

1. 基因间无义抑制突变 在编码蛋白质的基因中，无义突变的发生频率很高，例如密码子 AAG、CAG、GAG、UCG、UUG、UCG、UAC、UAU中任何一个密码子都可以通过一次碱

基替代突变而产生无义密码子 UAG，使肽链合成提前终止。如果这时能有一个突变的tRNA基因产物，其反密码子能够和 UAG 互补，则能将一个氨基酸安插在突变的无义密码子 UAG处，就能合成完整的蛋白质分子(图 5-3)。这个安插进来的氨基酸若就是野生型的氨基酸，当然就万事大吉；如果是别的氨基酸，蛋白质产物活性也可能部分恢复或全部恢复（也可能发生不能容许的氨基酸取代，因而产物并未恢复野生型产物的活性，例如赖氨酸取代了亮氨酸)。这种能抑制正向突变表型的突变了的 tRNA 称为抑制 tRNA(suppressor tRNA)。需要指出的是，抑制 tRNA 并不是由细胞对于无义突变应答反应的产物，而是自发突变或诱变的结果。

无义密码子有三种，UAG、UAA、UGA，因此无义突变也有三种，相应的无义抑制也就有三种。产生识别 UAG 的抑制 tRNA 的基因为琥珀型(amber)抑制突变，产生识别 UAA 的抑制 tRNA 的基因为赭石型(Ochre)抑制突变，产生识别 UGA 的抑制 tRNA 的基因为乳石型(Opal)抑制突变。这三种抑制突变均已在大肠杆菌中发现，用 *sup* 加上具体基因符号从 A 到 V 表示，其中研究较多的抑制 tRNA 基因列于表 5-2。

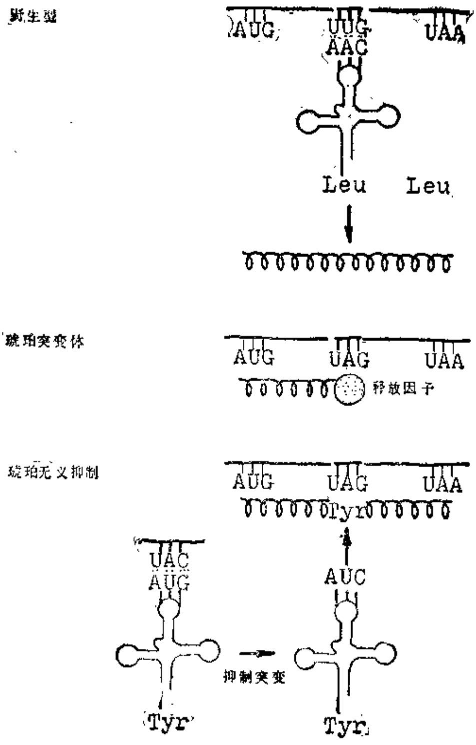



图 5-3 抑制 tRNA 对于无义突变的抑制

表 5-2 tRNA 基因的抑制突变

| 基因 | tRNA | 野生型 | | 抑制突变 | |
|---|---|---|---|---|---|
| | | 识别的密码子 | 反密码子 | 反密码子 | 识别的密码子 |
| *supD*(*su1*) | Ser | UCG | CGA | CUA | UAG |
| *supE*(*su2*) | Gln | CAG | CUG | CUA | UAG |
| *supF*(*su3*) | Tyr | $UA^{C}_{U}$ | GUA | CUA | UAG |
| *supC*(*su4*) | Tyr | $UA^{C}_{U}$ | GUA | UUA | $UA^{A}_{G}$ |
| *supG*(*su5*) | Lys | $AA^{A}_{G}$ | UUU | UUA | $UA^{A}_{G}$ |
| *supU*(*su7*) | Trp | UGG | CCA | UCA | $UG^{A}_{G}$ |

在三类抑制 tRNA 基因中，最容易鉴定的是 Amber 抑制突变，在 *E.coli* 中至少已鉴定出 6 种这样的 tRNA。它们均有反密码子 5′CUA3′。Ochre 抑制 tRNA 的反密码子为 UUA，由于反密码子和密码子识别过程中的碱基摆动原理，*E·coli* 所有识别 UAA 的 tRNA 也能识别 UAG。而在酵母中，Ochre 抑制 tRNA 只能识别 UAA，这可能是由于反密码子上碱基修饰的结果。

UGA 的抑制 tRNA (即 Opal 抑制 tRNA)是非常特别的。目前发现的 Opal 抑制tRNA 大多为 $tRNA^{Trp}$ 基因的突变产物。野生型 $tRNA^{Trp}$ 基因只有一个拷贝，反密码子为 CCA，识别唯一的色氨酸密码子 UGG。有一个 Opal 抑制 tRNA *supU*，其反密码子突变为 UCA。这个抑制 tRNA 不但能识别终止密码子 UGA，而且仍旧识别色氨酸的密码子 UGG。

另外一个UGA抑制 tRNA 为 *sup9*，它也是同样的 $tRNA^{Trp}$ 基因的突变产物。它的突变位点并不在反密码子上，而是在 DHU 臂的第 24 位上，由 G 突变为 A，使 GU 配对变为 AU 配对，从而增加了 DHU 双链结构的稳定性。然后，DHU 臂双链稳定性的增加又以某种机制改变了反密码子嘌呤的构型，进而改变 tRNA 在核糖体上的方向，使得反密码子 CCA 除了识别色氨酸密码子 UGG 外，还能与终止密码子 UGA 配对，其中反密码子中的 C 与密码子中的 A 为非同寻常的摇摆配对。尽管这一作用效率不高(故称为弱抑制 tRNA)，但是从无义突变的基因上产生少量有功能的产物，就足以使细胞存活下来。与此相一致的是在真核 tRNA 中的发现。牛肝中发现一种 $tRNA^{Ser}$，其反密码子为 mCCA，它并不识别色氨酸密码子 UGG，而是识别终止密码子 UGA。在此情况下，UGA 的终止作用受到部分抑制。

实际上，识别终止密码子 UGA 的能力可能并不是抑制 $tRNA^{Trp}$ 所特有的，很可能野生型的 $tRNA^{Trp}$ 就有识别 UGA 的能力，尽管这种识别效率很低。UGA 无义突变总是表现出一定程度的渗漏性，即使在缺少抑制 tRNA 时也是如此。这就说明野生型 $tRNA^{trp}$ 也能以低效率识别终止密码子 UGA。sup9 在 DHU 臂上的突变，与其说是新产生了识别 UGA 的能力，还不如说是提高了识别 UGA 的能力更为恰当。因此在大肠杆菌内，任何一个以 UGA 为终止密码子的基因，其蛋白质合成时有时能延长到下一个终止密码子。这就是所谓通读(readthrough)。通读还意味着，UGA 的终止作用的不完全可以作为在翻译水平上调控基因表达的一个手段。然而，这种 C 末端超长的蛋白质产物常常是没有功能的，有时对细胞是有害的。

一个细胞中，荷载某种氨基酸的 tRNA 反密码子发生突变，成为无义抑制 tRNA，那么该

种氨基酸的正常的密码子的识别是否受到影响呢？这是不会的。对于一种氨基酸甚至一种密码子来说，细胞中往往有两个或两个以上 tRNA 基因拷贝，这些基因在结构上并不完全相同。有的识别较常用的密码子，转录出的 tRNA 分子数目也多；有的只识别不常用的密码子，或者虽识别常用的密码子，但在细胞中的分子数少。后者就称为少数 tRNA(minor tRNA)。无义抑制突变和下面要说到的错义抑制突变大多发生在少数 tRNA 基因上。对于只有一个基因拷贝的 tRNA，其突变产物能同时识别无义密码子和原来的氨基酸密码子。因此，抑制突变的存在，并不影响正常的氨基酸密码子的识别。值得讨论的问题是在无义抑制 tRNA 存在时，细胞的正常终止功能是如何维持的？

上面已经说过，当细胞内含有 amber 抑制 tRNA 时，以单一的 UAG 为终止密码子的蛋白质模板可能被通读，而产生一个 C 末端过长的蛋白质产物，从而对细胞有害甚至引起细胞死亡。这个问题解决的机制是：

蛋白质合成的释放因子(RF 因子)与抑制 tRNA 对于终止密码子有强烈的竞争，因此抑制作用就减弱。假定 amber 抑制 tRNA 对无义突变的密码子识别效率为 50%，(注意：抑制效率约为10～50%，因为插入的氨基酸并不都能恢复产物活性)，则产生过长基因产物的几率也只有 50%，即正常终止的蛋白质产物仍有 50%。另一方面，细胞的蛋白质基因常采用双重终止密码子UAG～UAA。由于 Ochre 抑制 tRNA 很少，而且抑制效率很低，通常只有 1～5%，两个抑制效率的乘积只有 0.5%～2.5%(50%×1～5%)。UAG和UAA并不一定紧密相连，但中间间隔的碱基数必然是 3 的倍数，即处于同一个阅读框中。在 *E·coli* RNA 噬菌体 R17 和 MS2 中许多基因就是这种情况。

UAA 为很强的终止密码子，Ochre 抑制 tRNA 的抑制效率很低(1～5%)，UAA 是大肠杆菌中使用最多的终止密码子。Ochre 无义抑制不但效率低，而且具有 Ochre 无义抑制的细胞很不健康，其生长速度比正常细胞或只含 amber 无义抑制或Opal 无义抑制的细胞慢得多。这可能是因为大多数基因均以 UAA 为终止密码子，由于反密码子和密码子作用的碱基摇摆性，Ochre 抑制 tRNA 也同时能识别 UAG。这样，Ochre 抑制 tRNA 的存在，就影响到为数众多的基因表达的正常终止；尽管其识别效率很低，但可能某些对细胞活动十分关键的产物的表达量本来就低，Ochre 抑制 tRNA 就会造成严重的后果。换句话说，就是 UAA 的抑制突变具有强烈的选择负势和致死倾向，所以实际发生的频率低，效率低，生长速度慢。而UAG 和 UGA 单独作为终止密码子使用较少，能识别 UAG 和 UGA 的抑制 tRNA 的抑制效率就高得多。

2. 基因间错义抑制突变　错义突变的抑制机制与无义抑制很相似。假定正向的错义突变把GGA(Gly)变成 AGA(Arg)，则造成蛋白质失活，如果甘氨酸 tRNA 的反密码子也相应地由 UCC 变为 UCU，则能将甘氨酸安插在错义突变的密码子 AGA 处(图 5-4)，如果安插的氨基酸不是甘氨酸而是其他氨基酸，只要能恢复或部分恢复蛋白质活性，都表现出错义抑制效应。

除 tRNA 的反密码子的改变可以产生错义抑制以外，tRNA 结构其他部分的改变也可以导致错义抑制。这是由于无关的氨酰基-tRNA 合成酶与该 tRNA 作用使之带上了无关的氨基酸。而该 tRNA 对于密码子识别的特异性并没有改变。例如，*E·coli* 的 $tRNA_{I}^{Tyr}$氨基酸臂上6种点突变都引起谷酰胺酰-tRNA合成酶的错误识别而使 $tRNA_{I}^{Tyr}$ 错误地接载谷酰胺(图 5-5)。反密码子的改变，有时也会引起氨酰基-tRNA 合成酶的错误识别，该 tRNA 分子

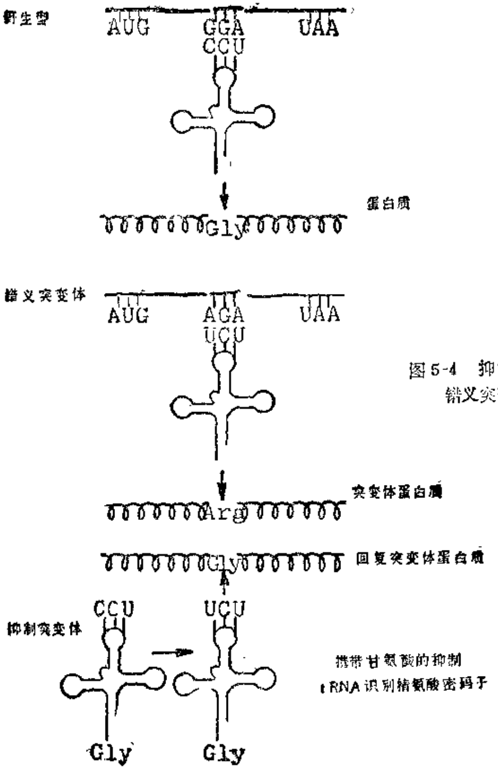



图 5-4 抑制 tRNA 对于错义突变的抑制

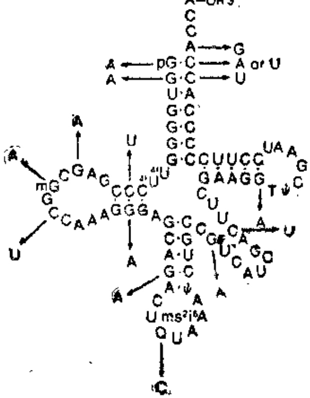



图 5-5 $tRNA_{I}^{Tyr}$ 的突变造成氨基酸的错误接载，箭头所指为突变后的碱基。

不但密码识别能力改变，而且接载的氨基酸也发生改变。这种突变 tRNA 可能为错义抑制 tRNA，也可能为无义抑制 tRNA。图 5-5 中 Q→C 即为后一种情况。其中 Q 为一种特别的修饰碱基 queuosine。这种情况下该 tRNA 不但成为 amber 抑制 tRNA，有时还错误地接载谷酰胺。侯雅明等人发现，丙氨酸 tRNA 分子上的 G3·U70 这一碱基对是丙氨酰-tRNA 合成酶的识别标记。如果将 $tRNA^{Cys/CUA}$ 和 $tRNA^{Phe/CUA}$ 这两个抑制 tRNA 的第 3 位和第 70 位换成 G 和 U，则它们不再荷载原来的半胱氨酸和苯丙氨酸，而是荷载丙氨酸。

氨酰基-tRNA 合成酶的突变也可造成氨基酸的错误接载，从而产生错义抑制。

需要指出的是，错义抑制必定是效率极低的，这是不难理解的。因为某个氨基酸错义成为 A 氨基酸，而每个基因上具有 A 氨基酸的密码子是很多很多的。若错义抑制效率高，则细胞几乎不可能有任何一种有功能的蛋白质。通常，错义抑制效率仅为 1%，个别的错义抑制效率很高，如在 AGA(Arg 密码子)处安插 Gly 的错义抑制有时可达 25% 的效率，这是因 AGA 这个精氨酸密码子很少使用，对正常的蛋白质造成的损害很小。

3. 基因间移框抑制突变　移框突变也可以由 tRNA 分子结构的改变而被抑制。例如 *E.coli* 的突变 $tRNA^{Gly}$ 在它的反密码子 loop 上增加了一个 C，反密码子成为 CCCC，因而能够识别 GGGG，从而矫正了由于插入一个碱基造成的移框突变(图 5-6)。一种突变的 $tRNA^{Lys}$ 可以识别 AAAU 和 AAAA，还有一种突变的 tRNA 能识别 ACCN，很可能这些突变的 tRNA 只与前三个碱基配对，而反密码子上新增加的碱基并不与密码子配对，而仅仅占据一个碱基的空间，使阅读框架向后移动一个碱基。

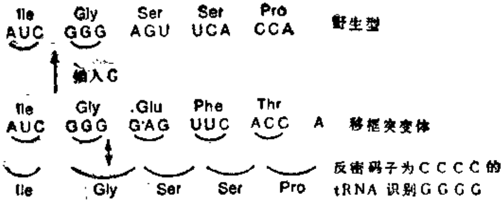


图 5-6　突变的 tRNA 对移框突变的抑制

由缺失点突变引起的移框也可以通过某些 tRNA 分子仅与两个碱基配对而得到矫正。这些 tRNA 分子可能是正常的 tRNA 分子，偶然地它们仅识别两个碱基，因此使所有的单碱基缺失突变总是呈现某种程度的渗漏。

4. 温度敏感抑制突变和抑制增强突变　许多 amber 突变的抑制 tRNA 表现出温度敏感性，即在 30℃ 能产生对 amber 突变的表型抑制，而在 42℃ 则不能产生这种效应。在这类抑制 tRNA 分子上，常常可以发现氨基酸臂碱基序列的改变，在 42℃ 时，tRNA 的分子形状发生很大变化，以致于影响到氨基酸的接载。人们对 Ts 抑制突变还没有进行详细的研究，然而它可能成为重要的实验手段。

如果细胞含有一个必需基因的 amber 突变，并含有 Ts amber 抑制 tRNA，在较高温时则生长很差，大多数形成很小的菌落，有时也出现长得很大的菌落。对这些大菌落细胞的分析表明，这些细胞含有一种次生突变，称为抑制增强突变(suppressor enhancing mutation)，写作 sue。sue 突变对于温度不敏感的抑制突变同样能增强抑制效应。现已发现三种不同的 sue 突变，称为 sue A、sue B、sue C，能增加 Amber 抑制效率 3～10 倍，人们对于抑制增强

突变的分子机制还不了解。

## 四、基因间间接抑制突变

基因间间接抑制的形式和类别多种多样，凡是能够使某一突变基因产物在一定程度上完成其应有使命的其他基因突变都是基因间间接抑制突变。如果原初突变发生在编码蛋白质的结构基因中，则造成间接抑制的机制主要有以下几种：

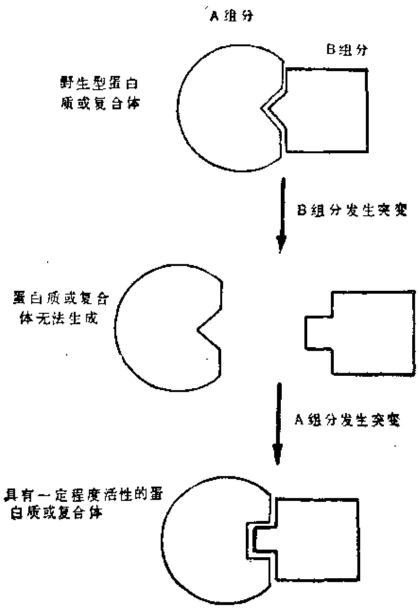


图 5-7 基因间间接抑制的途径之一

(1) 假设两个基因的多肽产物相互作用形成一个有功能的多亚基蛋白质复合体，其中一个基因的突变妨碍了正常相互作用。这时如果另一个基因的突变能够恢复这种相互作用。则后一个突变便是前一个突变的抑制突变(图 5-7)。这里，两个产物均为突变产物，细菌中这方面的例子很多。比如大肠杆菌 RNA 聚合酶 $\beta$- 亚基的某些突变能被 $\sigma$ 亚基的突变抑制；某些核糖体蛋白质的突变常常被另一些核糖体蛋白质的突变抑制。遗传学中常常通过选择这类抑制突变来分析蛋白质不同亚基或蛋白质复合体中不同组分之间结构上和功能上的关系。

(2) 在某一生化途径 A→B→C 中，负责催化 B→C 的反应步骤的 Y 基因发生突变后，该酶仅保存微弱的活性，所产生的 C 量太小而不足以使最终的表现型产生。如果催化 A→B 反应步骤的 X 基因产物因结构基因突变（提高 X 的活性）或基因的调节区域突变(增加产物 X 的浓度)。则中间产物 B 的浓度增加，在突变的 Y 基因产物作用下仍然能够产生较多的 C，从而产生应有的表现型。这里，X 基因突变就是 Y 基因突变的抑制突变。

(3) 当一个生化途径因突变被阻断时，抑制突变可以使反应绕过被阻断的步骤而补偿第一个突变的效应。例如由最初底物 A 产生最终产物 D 可以通过 A→B→D 和 A→C→D 两种途径，其中 A→B→D 为主要途径，A→C→D 为活性很低的支路。A→B→D 途径因突变失活后，细胞不能产生足量的 D；这时，任何提高 A→C→D 途径活性的突变都能增加 D 的产量，表现抑制效应。抑制突变还能够通过开辟一个新的生化途径而补偿第一个突变的效应，这方面研究得较多的例子有大肠杆菌的 *ebg* (evolved $\beta$-galatosidase，即进化的 $\beta$- 半乳糖苷酶）基因。正常情况下，大肠杆菌的半乳糖苷酶由 *lacZ* 编码，它的大片段缺失使酶失活，细胞不能利用乳糖作为唯一碳源。可是，从这种 *labZ* 缺失突变体中分离到了可以以乳糖为唯一碳源的菌株。这些菌株含有第二位点突变，位于 *ebg* 基因内，使原来无活性的 *ebg* 基因产物具有半乳糖苷酶的活力，*ebg* 基因和 *lacZ* 在染色体上相距甚远，它们的产物在分子量及其它特点上也

相差甚大，野生型细胞中 *obg* 基因的起源和功能尚不清楚。

(4) 有时一个基因的突变造成某种有害物质的积累，另一基因的突变可以通过提高分解酶的活性或将有害物质纳入其他生化反应的轨道而消除之。从表现型上，后一个突变也是前一个突变的抑制突变。

有时某基因的启动子、操作子、核糖体结合区域(S-D 序列)等调控区的突变妨碍了 RNA 聚合酶、阻遏蛋白或核糖体与它们的结合，使有关基因的转录或翻译的调控受到影响。这些突变的效应可以被突变的 RNA 聚合酶、阻遏蛋白或突变的核糖体组分所抑制。从效应上看，后一类突变也可以算作前一类的抑制突变。

综上所述，间接抑制突变基因和它所抑制的突变基因之间在结构上或功能上密切相关。在遗传学研究中，分离鉴定某突变基因的抑制突变体是鉴定各基因产物之间、基因产物与基因之间、基因产物与基因的调控位点之间相互关系的重要手段。

## 第四节　突变剂和突变生成[18~21]

### 一、碱基类似物在 DNA 复制时的渗入

除了四种标准的碱基 G、A、T、C 之外，有些与标准碱基相似的物质也能在 DNA 复制时渗入并与互补链上的碱基生成氢键而配对，从而抵抗 DNA 聚合酶的 3′→5′ 外切酶活性的校对作用。然而，如果仅是单纯的替代并不会引起突变，因为在下一轮 DNA 复制时又可以产生正常的 DNA 分子。然而，这些碱基类似物并不安分守已，而是朝秦暮楚，常常发生酮式和烯醇式的互变异构，在复制子代 DNA 时引起配对性质的改变，于是就造成了碱基替代突变，所有碱基类似物引起的替代都是转换而不是颠换。

5-溴尿嘧啶(BU)是一种常用的突变剂。在通常情况下，它以酮式结构存在，成为胸腺嘧啶的类似物，能与 A 配对，但它有时以较低的频率以烯醇式结构存在，它就不再与 A 配对，而是与 G 配对(图 5-8)。胸腺嘧啶也有酮式和烯醇式互变异构现象，但其烯醇式的发生率极低。而 BU 中由于 Br 是电负性很强的原子，因而其烯醇式结构的发生率要高得多。一般来说，BU 代替 T 渗入 DNA 的机会较多，产生 A·T→G·C 的转换；也可以代替 C 渗入 DNA，产生 G·C→A·T 的转换，不过后一能力没有前一能力高。不管哪种情况，BU 渗入 DNA 后必须经过两轮复制才能产生稳定的可遗传的突变(图 5-9)。

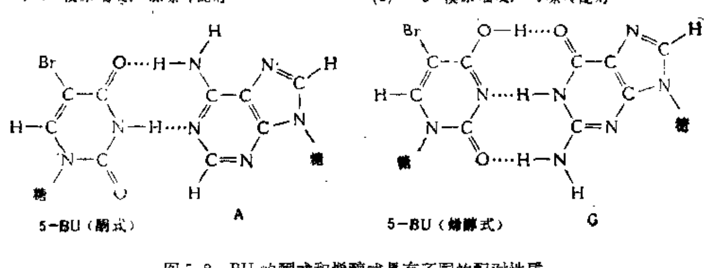



图 5-8　BU 的酮式和烯醇式具有不同的配对性质

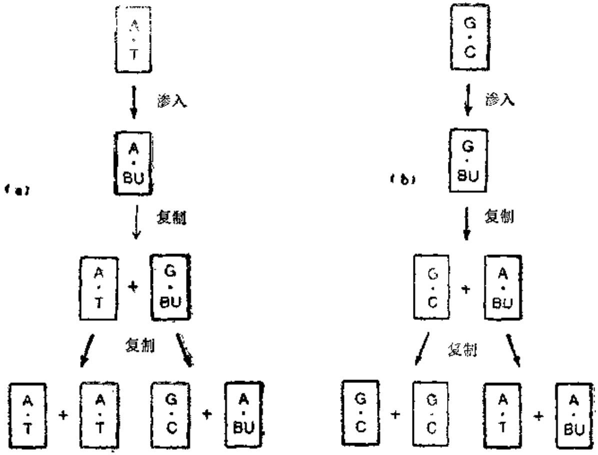
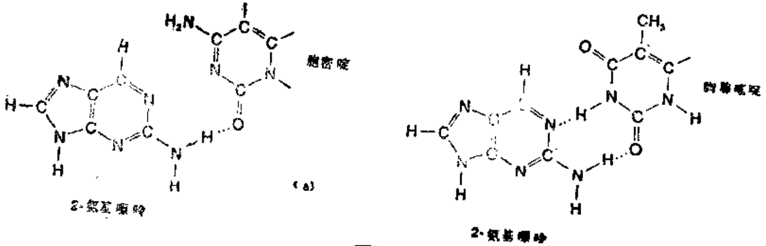



图 5-9　5-溴尿嘧啶引起的碱基替代突变

(a) BU 代替 T 而渗入 DNA ("复制错误")，产生 A·T→G·C 的转换

(b) BU 代替 C 而渗入 DNA ("渗入错误")，产生 G·C→A·T 的转换



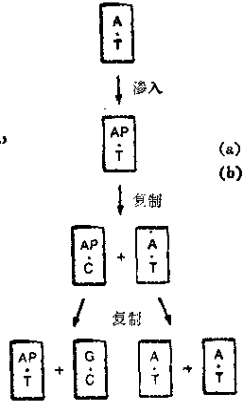




图　5-10

(a) 2-氨基嘌呤的不同配对性质

(b) 2-氨基嘌呤引起 A·T-G·C转换

另一种常用的碱基类似物突变剂是 2-氨基嘌呤(2-aminopurine，简称 AP 或 2-AP)。AP 与 BU 不同，AP 不能发生变构，但有不同的配对性质。一般情况下 AP 能代替腺嘌呤与胸腺嘧啶配对，但有时也能以一条氢键与胞嘧啶配对。由于 AP 与 C 的结合能力很弱，因此AP 主要产生 A·T→G·C 的转换，而一般不能产生 G·C→A·T 的转换(图 5-10)。

## 二、DNA 分子上碱基的化学修饰

许多化学物质都能以不同的方式修饰 DNA 的碱基，然后改变其配对性质而引起突变。最常见的化学突变剂有亚硝酸(nitrous acid 即 $HNO_2$)，羟氨(hydroxylamine 简称 HA)，甲磺酸乙酯 (ethyl mathanesulfonate 简称 EMS)， N-甲基-N′-硝基-N-亚硝基胍(N-methyl-N′-nitro-N-nitrosoguanidine，简称 NNG)，这几种突变剂的分子式示于图 5-11。

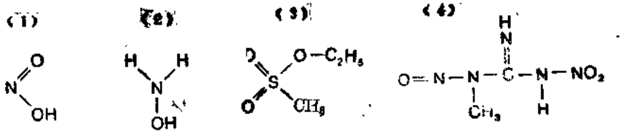


图 5-11 几种常见的化学突变剂

(1) 亚硝酸；(2) 羟胺；(3) EMS；(4) NNG。

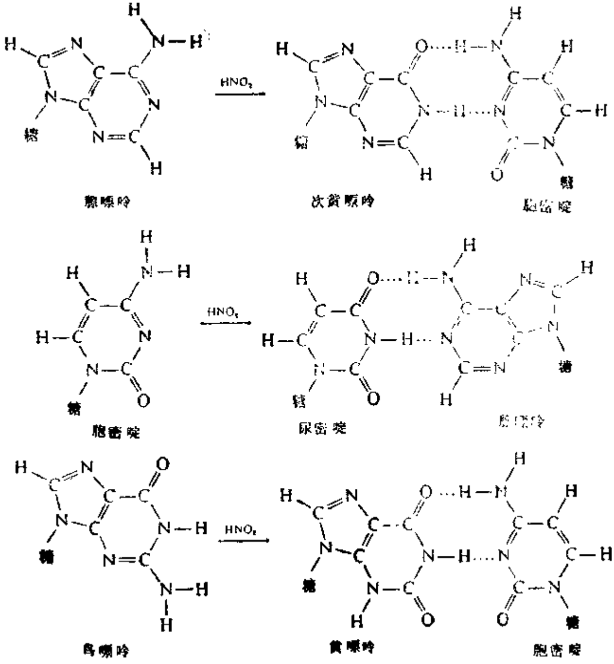
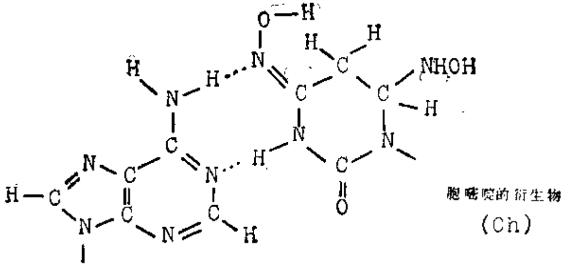



图 5-12 亚硝酸引起突变的机制

1. 亚硝酸引起的氧化脱氨反应 凡是含有 $NH_2$ 基的碱基(A、G、C)都可以被亚硝酸作用产生氧化脱氨反应，使氨基变为酮基，然后改变配对性质，造成碱基转换突变，与碱基类似物的突变机制相似，差别只在于碱基类似物是在DNA 复制时由外界渗入，而亚硝酸是氧化DNA链上已有的碱基。然后，也同样需要两轮复制才能产生稳定的突变。在亚硝酸作用下，胞嘧啶可以变为尿嘧啶，复制后可引起 G·C→A·T 转换；腺嘌呤可以变为次黄嘌呤(Hypoxanthine)，复制后可引起 A·T→G·C 的转换，鸟嘌呤可以变为黄嘌呤(xanthine)，它仍旧与 C 配对，因此不引起突变(图 5-12)。尿嘧啶，次黄嘌呤，黄嘌呤都可以为各自的糖基酶修复系统所修复；如果当修复系统还未来得及修复时 DNA 就开始复制，前二者则导致突变。

(a)




(b)

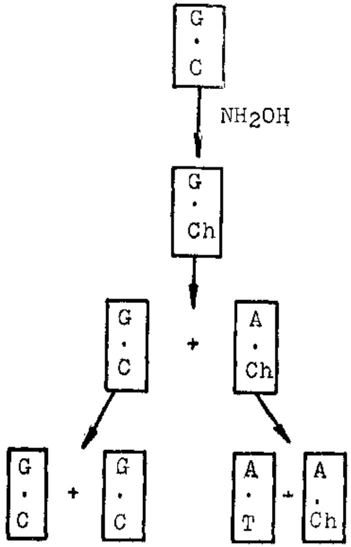



图 5-13 (a) 胞嘧啶与羟胺反应的衍生物与腺嘌呤配对 (b) 羟胺引起 G·C→A·T 转换

2. $NH_2OH$ 的致突变作用　$NH_2OH$ 的作用机制很复杂，在不同的 pH 和不同的 $NH_2OH$ 浓度时有不同的产物。在 pH6.0，浓度 0.1～1.0 mol/L 时，$NH_2OH$ 与胞嘧啶起反应，生成的产物能与腺嘌呤配对，从而导致 G·C→A·T 转换(图 5-13)。

3. 烷基化试剂的致突变作用　常用的烷基化试剂有硫酸二甲酯，甲磺酸乙酯(EMS)，乙磺酸乙酯等。烷基化位点主要在鸟嘌呤的 N-7 位上和腺嘌呤N-3位上，在腺嘌呤的N-7位有时也可以烷基化。由于鸟嘌呤上 N-7 的烷基化，使之成为带一个正电荷的季铵基团，这个季铵基团产生两个效应：一是促进第 1 位氮基上氢的解离，使 G 不再与 C 配对而是与 T 配对，从而造成G·C→A·T转换。N-7成为季铵基团后，减弱了N9位上的N-糖苷键，而产生了去嘌呤作用。大部分的无嘌呤位点都可以被无嘌呤内切酶系统所修复，但是有时复制在修复之前进行，则在无碱基的位置上可以插入任何一个碱基。在第二轮复制以后，则原来的 G·C 对可以变为任何碱基对 G·C，C·G，A·T，T·A，既有转换，又有颠换(图5-14)。用较低的 pH 或较高的温度或二者结合去处理噬菌体 DNA，则可产生去嘌呤作用。这样处理过的噬菌体DNA在复制时可产生大量的颠换突变，这与烷化剂的作用机制是一致的。

NNG 也是一种烷化剂，能使碱基烷基化。NNG 诱变能力很强，它可以使一个群体中任何一个基因的突变率高达 1%。NNG 处理产生的突变体常常含有多位点突变，而且这些突变位点往往在同一个基因内或相邻基因内成簇分布，这大概是由于 NNG 对复制叉附近的碱基序列诱变能力最强。利用 NNG 进行诱变的优点是很容易获得突变，而缺点也恰恰是突变体携带了许多不需要的突变，影响了对突变的分析。

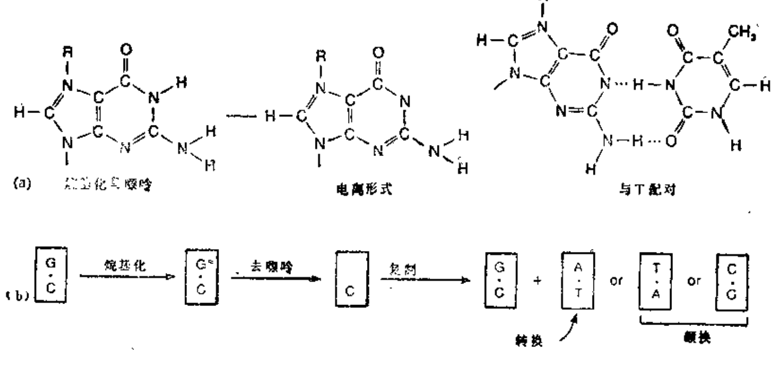



图 5-14　(a) N-7 烷基化鸟嘌呤由于 N-1 位上 H 的电离而与胸腺嘧啶配对，从而导致G·C→A·T 转换　(b) N-7 烷基化鸟嘌呤通过去嘌呤作用导致转换和颠换

## 三、嵌合剂的致突变作用

吖啶橙(acridine)，原黄素(proflavine)，吖黄素(acriflavine)等吖啶类染料分子均含有吖啶稠环。这种三环分子的大小与 DNA 的碱基对大小差不多，可以嵌合到 DNA 的碱基对之间，于是原来相邻的两个碱基对分开一定的距离，含有这种染料分子的 DNA 在复制时，由于某种目前尚不知晓的原因，可以插入一个碱基，偶尔也有两个。这样就出现一个或几个碱基对的插入突变。有时也有很低频率的单碱基缺失突变。所以这些突变，都引起阅读框的改变

(图5-15)。

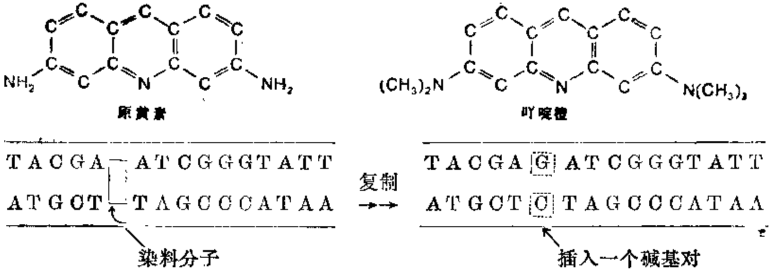



图 5-15　嵌合剂引起插入突变和移框突变

### 四、转座成分的致突变作用

生物体内含有许多转座成分，它们一般长数百至数千 bp，可以通过一种复杂的转座机制将其一个复制拷贝插入到基因组的另外一个位点。如果这个位点处于一个基因内部，这种大片断的插入常常造成基因失活。*E. coli* 的 Mu 噬菌体与转座成分很类似，它可以整合到 *E. coli* 染色体的任何位置，从而引起突变，故名 Mu 噬菌体(*mutator* phage)。

### 五、增变基因

生物体内有些基因与整个基因组的突变率直接相关。当这些基因突变时，整个基因组的突变率明显上升，因此把这些基因称为"增变基因"(mutator genes)。实际上这是一种误称。因为这些基因在野生状态时恰恰是维持 DNA 圣洁性的因素，只是由于其自身难保才犯渎职之过。在所谓增变基因中目前已经了解的有两类。一类是 DNA 聚合酶的各个基因。如果 DNA 聚合酶的 3′→5′ 校对功能丧失或降低，则使突变率升高而且随机分布。另一类是 *dam* 基因和 *mut* 基因，如果这些基因突变，则错配修复系统功能丧失，则引起突变率升高。

### 六、紫外线的致突变作用

紫外线照射 DNA，由于高能粒子的直接作用和自由基的间接作用，引起了 DNA 分子的各种损伤，嘧啶二聚体是其中一种重要的损伤。这些损伤，诱发了错误潜伏的 SOS 修复系统，从而导致很高的突变率。

紫外线引起的突变包括各种形式的转换和颠换。突变不是完全随机的，G·C→A·T 转换最多，而且有某种序列优先性，如 C 处在 TCA 序列中产生的 C→T 转换率要比在 ACA 序列中高 15 倍。除碱基替代突变外，紫外线还能引起缺失、重复和移框突变。这些突变可能是直接作用，间接作用和 SOS 系统共同作用的结果。

为研究 SOS 修复系统本身的诱变作用，有些实验以非诱变剂处理的方法触发 SOS 反应。例如，rif-1 的突变改变了 RecA 蛋白，能使细胞出现 SOS 反应。在这种情况下，各种转换和颠换也同样发生，但以 G·C→T·A 颠换居多，而且也具有序列优先性，如 CGA 序列中的 G→T 颠换率是 AGA 或 GGA 序列中的 20 倍以上。已知在 *E·coli* 中，SOS 系统的诱变作用与 *umuC* 基因关系十分密切。*umuC* 的突变使 SOS 系统的诱变作用丧失，因此 *umuC* 基因产物很可能是错误潜伏修复系统的重要组分。目前尚未分离到 *umuC* 基因产物。SOS 系统的诱

变作用的另一种可能途径是 RecA 蛋白促进了不完全同源的 DNA 序列之间的重组，产生了大量的错配碱基，从而引起突变。

除了紫外线外，各种电离辐射，如 X 射线，$\gamma$ 射线以及丝裂霉素 C (mitomycine)、黄曲霉素 $B_1$(aflatoxin $B_1$) 等致癌物质都能诱发 SOS 反应。

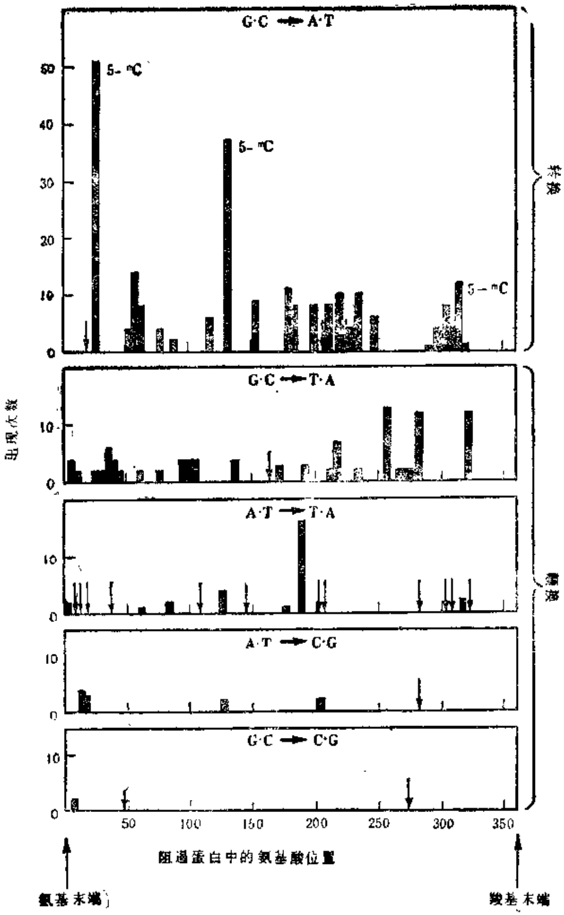



图 5-16　乳糖操纵元阻遏蛋白基因中自发的无义突变的分布和性质。图中显示出 G·C→A·T 转换和全部四种颠换。竖直箭头表示可能出现无义突变之处，但实际上并未观察到。在转换一图中，容易看出，5-mC 是产生突变热点的重要原因

### 七、突变热点

从理论上讲，DNA 分子上每一个碱基都能发生突变，但在实际上突变位点并非完全随机分布。DNA 分子上的各个部分有着不同的突变频率。Seymour Benzer 采用各种突变剂处理 T4 噬菌体，选择了大约 1 500 个 rⅡ基因的突变体，然后用遗传学方法分析突变在基因的不同位点上的分布情况。结果发现它们分布在大约 500 个位点上，可是它们并不遵循随机碰撞事件的泊松分布规律(poisson distribution)，而是在某些位点的频率大大高于平均数(图(5-16)。这些位点就称为突变热点(hot spots of mutation)。不仅在诱发突变中有突变热点，而且在自发突变中也有突变热点。

突变热点长期以来一直是一个谜。通过分子遗传学的研究，现在已经有了比较清楚的认识。形成突变热点的最主要的原因是 5-甲基胞嘧啶(MeC)的存在。关于碱基甲基化的功能前面已经讲过。MeC 和 C 一样，在突变剂的作用之下，会产生脱氨氧化。所不同的是，C 脱氨氧化后生成 U，U 可以为尿嘧啶糖基酶系统所修复，突变的频率就很低；而 MeC 脱氨氧化后生成 T，而 T 是 DNA 的正常组分，形成 G·T 的不配对状态。这种作用如果发生在 DNA 复制期间的新生链上，就可以被错配修复系统所修复。然而，这种作用也可以发生在 DNA 非复制期。这时，由于两条链的甲基化程度相同，错配修复系统就失去了判别标准，只能随机地切除一个，就有一半的可能性发生突变。如若发生在正在复制的模板链上，则很快引起突变。用亚硝酸作突变剂，在 5-甲基胞嘧啶处，明显地出现突变热点，就是这种道理。

形成突变热点还有其他原因。在短的连续重复序列处容易发生插入或缺失突变，插入或缺失的正是这一重复序列，这是由于在 DNA 复制时发生模板链和新生链之间碱基配对的滑动而造成的。例如 *E·coli lacI* 基因中有三个连续的 *CTGG* 序列，很容易产生一个 *CTGG* 序列的插入突变或缺失突变。

突变热点还与突变剂有关，使用不同突变剂时出现的突变热点也不同。这是容易理解的，因为不同的突变剂作用机制不同，有的 DNA 序列使某个碱基对突变剂更为敏感，有的则相反。例如用 5-溴尿嘧啶处理 λ 噬菌体的 cI 基因，ACGC 序列中 A→G 转换率比 A—C—非G—N 序列中的 A 高 15 倍，比 A—非 C—N—N 序列中的 A 高 100 倍。转座因子的致突变作用，增变基因的作用，以及紫外线的作用，也都具有某种不同程度的序列优先性。

## 第五节　离体定向诱变[18,21~30]

前面所讲的自发突变和诱发突变都是在细胞内发生的。这些突变虽然能表现出一定的位点特异性，并且根据诱变剂的种类可以预先知道突变的类型(如转换、颠换、移框等)，但还不能按照预先设计的核苷酸序列定向制造突变，而只能先从表现型上的变化发现突变，继而分离突变体，再采用遗传学方法和其他方法进行分析，才能知道突变的性质。随着重组 DNA 技术、DNA 序列分析技术，寡聚核苷酸合成技术以及其他分子生物学技术的出现和迅速发展，现在人们已经能有目的地在离体条件下制造位点特异性突变，然后再设法导入生物体内，观察并分析这些突变的表现型。这种在体外条件下定向诱发突变的方法，我们称之为离体定向诱变。用离体定向诱变技术来制造和分析突变，在方法上与经典遗传学方法相反，因此又常被称为“反向遗传学”(reverse genetics)或 DNA 分析遗传学(genetics by DNA analysis)。目前

人们已经可以定向地制造缺失、插入、碱基替代、移框等不同类型的突变。这些方法已在研究基因的表达和调控机制方面作出了很大贡献，将来在分子遗传学和生物化学的各个研究领域，特别是蛋白质工程中必然发挥更大的作用。利用各种核酸酶和突变剂以及其他试剂能设计出许多制造突变的办法，但限于篇幅，这里不能作详细介绍。下面对于利用寡聚核苷酸进行离体定向诱变的方法加以扼要的叙述。这种方法通常称为定点突变(site-directed mutagenesis)。

利用人工合成的寡聚核苷酸可以制造任何部位的位点特异性突变，而不受限制酶切点的限制。如果希望改变某个 DNA 克隆的某一个特定碱基，可以先合成一条包括靶碱基及其附近序列的寡聚核苷酸，一般长约 15～20 碱基(靶碱基位置处于中间)；如果需代换的碱基多，寡

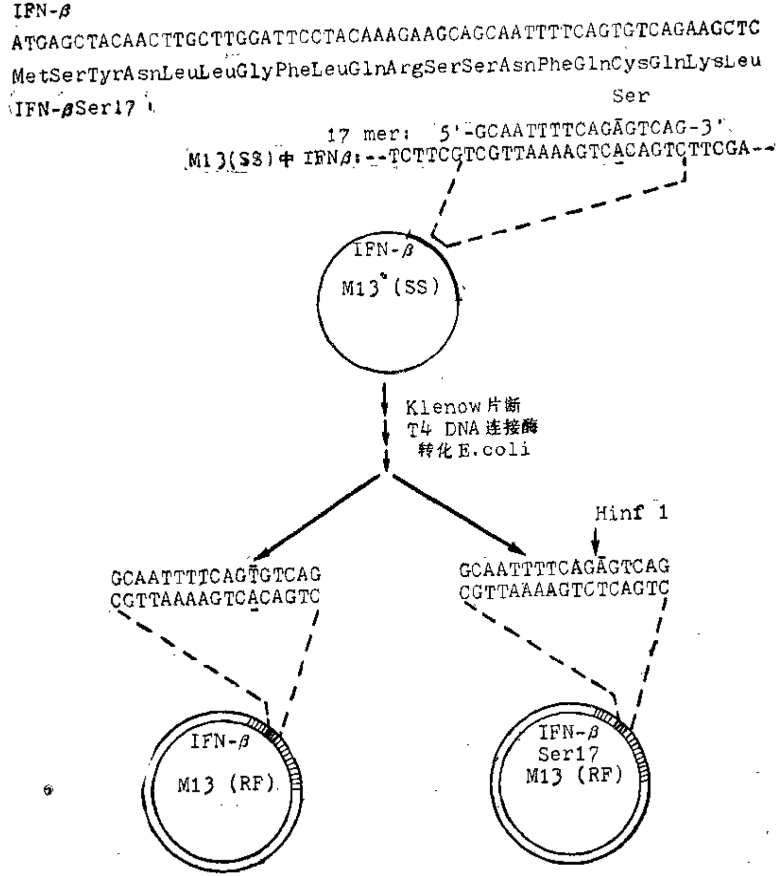



图 5-17 利用人工合成寡聚核苷酸产生 IFN-β 基因的碱基替代型突变。
左图：野生型 IFN-β 克隆；右图：突变型 IFN-βSer17 克隆。

聚核苷酸的长度还要长一些。这条寡聚核苷酸链除了要替换的靶碱基外，其余的序列与野生型 DNA 分子中的相应序列完全一致。然后将合成的寡聚核苷酸和由单链噬菌体作载体（如 M13)所携带的 DNA 克隆的互补单链混合，进行分子杂交。由于存在着一个或几个非配对碱基，杂交反应必须在较低温度下进行。这段配对的寡聚核苷酸可以作引物，在 DNA 聚合酶 I Klenow 片断的作用下合成完整的互补链，再用 DNA 连接酶连接起来。将此双链 DNA 导入大肠杆菌中，经修复和 DNA 复制就可以得到可以稳定遗传的突变 DNA 克隆(图 5-17)。从表面上看，复制后产生的子代 DNA 克隆中应有一半为突变体，但实际上要少得多，大约只有 10～15%。这可能有如下的原因：一是杂交分子中原来的单链的甲基化程度高，在导入大肠杆菌后，错配修复系统倾向于以原来单链为模板修复不配对碱基，因而切除了寡聚核苷酸上的错配碱基；二是在体外合成互补链时，DNA 聚合酶在合成一圈互补链后继续以滚环复制方式合成 DNA，从而把寡聚核苷酸置换出来。用 T 4 DNA 聚合酶代替 Klenow 片断可以得到较好的效果。图 5-17 表示要将干扰素 $\beta$ 的第 17 个氨基酸由Cys 变为Ser，Cys 的密码子为 TGT，而 Ser 的密码子为 AGT，因此只要改变一个碱基对就够了(图 5-17A)。然后人工合成一条 17 个碱基长的寡聚核苷酸与已经克隆在 M13 中的干扰素 $\beta$ 的反义链互补，并以这寡聚核苷酸为引物进行互补链的合成。然后转化大肠杆菌，经过大肠杆菌错配修复酶系的修复然后复制或直接进行复制。则产生野生型和突变型两种基因克隆(图 (5-17B)。由于突变基因引入了一个限制性内切酶 Hinf1 的切点，因此突变体很容易被筛选出来。同样的道理，如果突变体基因破坏了原来野生型基因的限制酶位点，也可以作筛选的指标。

如果碱基替代没有改变或产生可以利用的限制酶位点，则可使用杂交方法。上面已说到，由于寡聚核苷酸与目的基因杂交时由于不配对碱基的存在，只有在低温时才能产生杂交分子。这样，将待检的 DNA 克隆放在硝酸纤维素滤膜上，溶菌，将 DNA 变性，然后用放射性 $^{32}P$ 标记的寡聚核苷酸(与诱变时使用的相同，差别只是有放射性标记)作探针来进行杂交。可以选择不同的温度进行。在较低温度时，探针可以和突变型分子杂交，也能和野生型分子杂交；在较高温度时，探针与两种分子都不能杂交。在这两种温度之间必然有一种临界温度，这时探针只能与突变型分子杂交而不能与野生型分子杂交，通过放射自显影，就可以判断哪些克隆是突变体，哪些是野生型。

产生了突变的 DNA 分子以后，就可以通过等位基因代换的方法将突变基因送回细胞内的原来的基因组中。只有这样，才能在生理条件下检验突变效应；从改造物种的角度来说，这才算完成了定向诱变的全过程。如果所研究的基因组很小，如病毒、噬菌体、质粒的 DNA 等，方法很简单，用重组 DNA 技术就可以把突变基因送回原基因组。比如，某病毒 DNA 可以被限制酶 HindⅢ 和 BamH1 切成两段，将其中一段克隆在载体分子中进行定向诱变处理（载体也恰恰具有这两个限制酶的切点或事先加上人工接头)，然后再用同样的限制酶将诱变好的片断切下来，用 DNA 连接酶将它与未经处理的片断连接起来，即得到完整的含有定向突变的病毒 DNA 分子。

将定向突变的基因送回细胞染色体中去的情况就复杂得多。对于可转化的细菌和酵母细胞，可以采用 DNA 同源重组的办法以突变的基因取代野生型的等位基因。要经过两次交换事件，突变基因才能取代野生型等位基因。这里有两种选择突变体的方法，一种是直接选择两次交换发生后所产生的转化体，这实际上就是经典的细菌转化方法。另一种方法是分步选择法，即在每一次交换后都作一次选择。对于没有指标可供选择的突变基因，往往在制作 DNA

克隆时把一个可被鉴别的标记(如抗药性基因，原养型基因等)与突变基因结合在同一个重组DNA分子中。然后通过分步选择并促进突变基因的取代。图5-18即为分步选择法的一个例子，该实验的最终目的是将离体定向诱变的酵母 *his*3⁻ 基因(小片断缺失)取代染色体上野生型 *his*3⁺ 基因。*his*3 基因是负责合成组氨酸的基因之一。用负责尿嘧啶合成的基因 *ura*3作为标记与 *his*3⁻ 克隆在一起。选择 Ura3⁻(基因型为 *his*3⁺ *ura*3⁻)作为转化受体。转化以后，首先选择具有 Ura⁺ 表现型的转化体又同时是 His⁺ 表现型，然后将这些转化体长时间放在含有组氨酸和尿嘧啶的培养基上生长，这样就撤除了野生型基因 *his*3⁺ 和 *ura*3⁺ 的正向选择压力，原来整合进染色体的DNA克隆又可能通过一次同源交换而被切除。有时切除的并不是原来克隆上的 *his*3⁻ 和 *ura*3⁺，而是 *his*3⁺、*ura*3⁺。这样，染色体中就留下了定向突变的 *his*3⁻ 基因。通过选择 Ura⁻ His⁻表现型个体就可以获得理想的含有 *his*3⁻基因的突变体。到目前为止，真核生物核基因的代换只在酵母中有成功的报道，对于高等动植物的核基因的代换

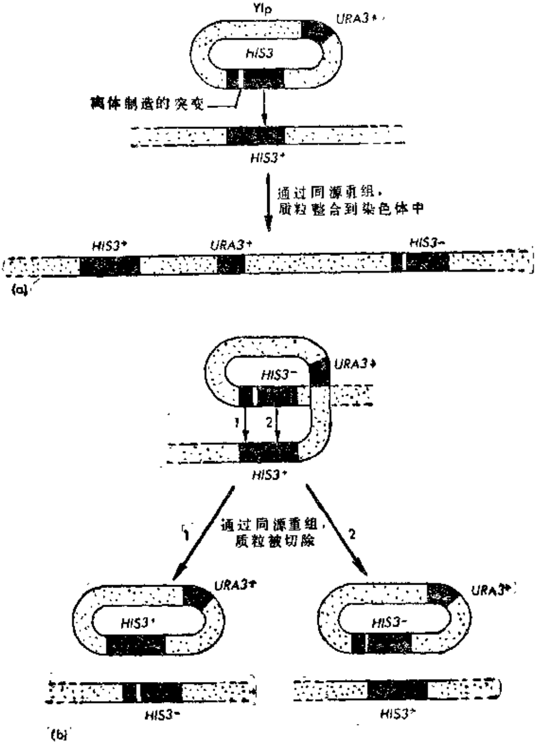



图5-18酵母突变型基因 *his*3⁻ 取代野生型基因 *his*⁺ 的过程

(a) 带有 *His*3⁻ 基因的质粒与染色体整合；(b)质粒的切除，途径1将 *His* 3⁻ 基因留在染色体上，而途径2则恢复了野生型染色体。

还有很大的困难。对于高等植物，人们已经将外源基因引入基因组。例如1986年美国加州大学D.R.Helinski已将萤火虫的萤光素酶基因通过植物病毒和Ti质粒插入到烟草的基因组中，从而“点亮”了烟草。但这还不是基因代换。高等植物的基因工程之所以能走在高等动物之前，主要是因为植物的一个细胞能够长成为植株，只要用培养的植物细胞进行基因操作就行了。而就目前所知，高等动物细胞则不具备这种特性，可能需要开拓新的途径。据报道，1989年意大利生物医学技术研究所和中国科学院动物研究所先后用精子作为外源基因的载体，成功地培育出转基因动物。这是动物基因工程的重大突破。

## 提要

突变是一种遗传状态，是可以通过复制而遗传的DNA结构的永久性改变。携带突变的生物个体或群体或株系叫做突变体。野生型是指有机体的正性状，如分解某种底物的能力或合成某种物质的能力，而突变体一般缺乏这种能力或者这种能力较差。自然界的因素作用于有机体，产生自发突变，其频率很低；而诱发突变的频率较高；在实验室中还可以进行定点突变。条件型突变在研究工作中具有重要意义。

能够抵消正向突变效应的突变称为回复突变。真正的原点回复突变很少，大多数为第二点回复突变。第二点回复突变可以发生在正向突变的基因内，也可发生在其它基因之中。无义抑制tRNA，错义抑制tRNA和移框抑制tRNA是重要的基因间抑制突变。基因间间接抑制突变的花样繁多，代偿性突变中不少是属于这类突变。

自然界存在着许多突变剂。碱基类似物引起的突变均为转换。亚硝酸盐、羟氨、烷基化试剂，吖啶类稠环化合物等都有很强的致突变作用。转座成分常造成插入突变。DNA复制系统和错配修复系统基因的突变则导致整个基因组突变率的显著上升。紫外线可通过其直接作用以及诱发的SOS修复导致突变。DNA分子的不同序列具有不同的突变几率。5-甲基胞嘧啶是造成突变热点的重要原因。

利用噬菌体M13的单链DNA为载体，以寡聚核苷酸为引物进行定点突变的方法是一种重要的离体定向诱变方法，是基因工程的重要手段。利用同源重组的原理可以进行野生型基因的代换，但人们仍然需要开发更为有效的基因转移途径。

## 参考文献


1. Miller, J. H. (1983) "Mutational Specificity in Bacteria." *Ann. Rev. Genetics*, 17: 215~238.
2. Paterson, M. C. and Gentner, N. E. (1984) "Introduction: Environmentally Induced DNA Lesions and Their Biological Consequences." in *Repairable Lesions In Microorganisms*, ed. Hurst, A. and Nasim, A., Academic pres. New York, pp.1~56
3. Miller, J. H. and Low, K. B. (1984) "Specificity of Mutagenesis Resulting from the Induction of the SOS System in the Absence of Mutagenic Treatment." *Cell* 37: 675~682.
4. Helinski, D. R. and Yanofsky, C. (1963)" A Genetic and Biochemical Analysis of Second Site Reversion." *J. Biol. Chem*, 238: 1043~1048.
5. Brenner, S. et al. (1965) "Genetic Code: The Nonsense Triplets for Chain Termination and Their Suppression". *Nature* 206: 994~998.
6. Hill, C. W. (1975) "Informational Suppression of Missence Mutations." *Cell* 6: 419~427
7. Steege, D. A. and Soll, D. G. (1979) "Suppression." in *Biological Regulation and Development I*, ed. Goldberger, R. F. Plenum Press, New York, pp. 433~486.
8. Goodman, H. M. et al. (1968) "Amber-Suppression: A Nucleotide Change in the Anticodon of a Tyrosine Transfer RNA." *Nature* 217: 1019~1024

9. Hirsh, (1971) "Tryptophan Transfer RNA as the UGA Suppressor," *J. Mol. Biol.* 58: 439～458.
10. Ozeki, H. et al. (1980) "Genetics of Nonsense Suppressor tRNAs in *Escherichia coli.*" in *Transfer RNA, Biological Aspects*, ed. Soll, D. et al., Cold Spring Harbor Laboratory, New York, pp. 341～349.
11. Weiner, A. M. and Weber, K. (1973) "A Single UGA Codon Functions as a Natural Termination Signal in the Coliphage Qβ Coat Protein Cistron." *J. Mol. Biol.* 80: 837～855.
12. Carbon, J. et al. (1966) "Missense Suppression Due to a Genetically Altered tRNA." *Cold Spring Harbor Symp. Quant. Biol.* 31: 487～497.
13. Hou, Y. -M. and Schimmel, P. (1988) "A Simple Structural Feature Is a Major Determinant of the Identity of a Transfer RNA." *Nature* 333: 140～145.
14. Roth, J. R. (1981) "Frameshift Suppression." *Cell* 24: 601～602.
15. Weiss, R. (1984) "Molecular Model of Ribosome Frameshifting." *Proc. Nat Acad. Sci.* 81: 5797～5801
16. Weiss, R. B. et al. (1988) "Reading Frame Switch Caused by Base-Pair Formation between the 3'End of 16S rRNA and the mRNA during Elongation of Protein Synthesis in *Escherichia coli.*" *The EMBO J.* 7: 1503～1507.
17. Gold, L. (1988) "Posttranscriptional Regulatory Mechanisms in *Escherichia coli.*" *Ann. Rev. Biochem.* 57: 199～233.
18. Smith, M. (1985) "In vitro Mutagenesis." *Ann. Rev. Genetics* 19: 423～463.
19. Hurst, A. and Nasim, A., ed. (1984) *Repairrable Lesions in Microorganisms*, Academic Press, New York.
20. Shortle, D. and Nathans, D. (1979) "Regulatory Mutants of Simian Virus 40: Constructed Mutants with Base Substitutions at the Origin of Viral DNA Replication" *J. Mol. Biol.* 131: 801～817.
21. Shortle, D. et al. (1981) "Directed Mutagenesis." *Ann. Rev. Genetics* 15: 265～294.
22. Winter, G. et al. (1982) "Redesigning Enzyme Structure by Site-Directed Mutagenesis." *Nature* 299: 756～758.
23. Zakour, R. A. and Loeb, L. A. (1982) "Site-Specific Mutagenesis by Error-Directed DNA Synthesis." *Nature* 295: 708～710
24. Heffron, F. et al. (1978) "In vitro Mutagenesis of a Circular DNA Molecule by Using Synthetic Restriction Sites." *Proc. Nat. Acad. Sci.* 75: 6012～6016.
25. Zoller, M. J. et al. (1982) "Oligonucleotide-Directed Mutagenesis Using M13-Derived Vectors: an Efficient and General Procedure for the Production of Point Mutations in Any Fragment." *Nucl. Acids Res.* 10: 6487～6500.
26. Zoller, M. J. et al. (1983) Oligonucleotide-Directed Mutagenesis of DNA Fragments Cloned into M13 Vectors. *Meth. Enzymol.* 100: 468～500.
27. Kunkel, T. A. (1985) Rapid and Efficient Site-Specific Mutagenesis without Phenotypic Selection. *Proc. Nat. Acad. Sci.* 82. 488～492.
28. Wood, W. et al. (1985) Base Composition Independent Hybridization in Tetramethylammonium Chloride: a Method for Oligonucleotide Screening of Highly Complex Gene Libraries. *Proc. Nat. Acad. Sci.* 82: 1585～1588.
29. Scherer, S. and Davies, R. W. (1979) "Replacement of Chromosome Segments with Altered DNA Sequence Constructed *in vitro*", *Proc. Nat. Acad. Sci.* 76: 4951～4955.
30. Rothstein, R. J. (1983) "One-Step Gene Disruption in Yeast." *Meth. Enzymol.* 101. 202～211.

# 第六章 转 录

## 第一节 概 述[1~4]

DNA 分子所储藏的蛋白质遗传信息，必须转变成信使 RNA 分子(mRNA)，才能到达蛋白质合成的工厂——核糖体，然后，蛋白质合成酶系才能把 mRNA 所带来的信息翻译成蛋白质。在合成蛋白质过程中，还必须有能携带氨基酸的 tRNA 和构成核糖体的 rRNA。这三类 RNA 都必须以 DNA 为模板，在依赖于 DNA 的 RNA 聚合酶催化之下合成，包括 RNA 链的起始、延伸、终止等步骤。这一系列过程就叫做转录。在由 DNA→RNA→蛋白质这样一个信息流中，RNA 是个中心环节。这样的信息流体制是何时建立起来的以及怎样建立起来的，现在无法说清楚这个问题。据认为，在原始的生命体中，遗传信息的携带者很可能是 RNA，它同时又是翻译蛋白质的模板。由于遗传信息需要有高度的稳定性，因而在绝大部分的生命体内，DNA 取代了 RNA 而成为遗传信息的携带者。特别是真核生物中由核膜将遗传信息的载体与蛋白质合成场所分隔开来，给细胞带来巨大的好处，使真核生物更能适应自然环境。在生物体内，转录并不是产生 RNA 的唯一途径。许多 RNA 病毒可以以其 RNA 为模板复制新的 RNA (即 RNA→RNA)。在还原病毒中还能以 RNA 为模板进行反转录从而产生互补的 DNA 链(即 RNA→DNA)。

转录是基因表达的第一步，也是最关键的一步。生物体基因表达调控的第一步也就是决定是否要让该基因转录。对于大多数基因来说，这是最重要的调控机制；在有些情况下甚至是唯一的调控机制。这是不难理解的。对于任何一个链锁过程，控制其第一步是最为有效和最为经济的。

论述转录过程，要涉及到两个方面。一是 RNA 合成的酶学过程，二是 RNA 合成的起始信号和终止信号，也就是 DNA 分子上的特定序列。人们在 50 年代末就开始了 RNA 合成的酶学过程的研究，了解得也比较清楚；而对 DNA 分子上的特别序列的研究，还是近 10 多年的事，因为只有在 DNA 重组技术和序列分析技术发展之后才有可能进行研究。关于 DNA 和 RNA 的碱基序列，方向总是从 5′→3′，除非特别说明。对于 DNA，一般总是讲非模板链的序列。文献上对转录时 DNA 两条链的命名稍微有点混乱。一般在 60 年代的文献上，常把作为转录模板的那条链称为有义链(sense strand)。而较新的文献则相反，把不作模板转录的链称为有义链，又称编码链(coding strand)；而把作为模板转录的链称为反义链(antisense strand)或模板链(template strand)。本书一律采用这种新的表示方法，选用非模板链的序列作为有义链，其方向和核苷酸序列都与这一区域转录出来的 RNA 序列相一致(只是 T 和 U 的区别而已)。对于 mRNA，也容易将 DNA 序列直接与氨基酸的密码子联系起来。这样，单链 DNA 噬菌体，如 $\phi\times174$，其 DNA 进入寄主细胞后必须复制出一条互补链而成为双链超螺旋的形式，才能从这条互补链上转录出它所需要的 RNA。这样，这条互补链就称为反义链，又称为负链；而能够进入噬菌体颗粒的 DNA 链则称为有义链，又称为正链。对于单链 RNA 病毒，若其基因组 RNA 能够作为 mRNA 或与 mRNA 相同，则该 RNA 病毒就称为正链 RNA 病

毒。若基因组 RNA 进入寄主细胞后必须通过 RNA 复制，产生出互补链作为 mRNA，则这种 RNA 病毒就称为负链 RNA 病毒。这是 RNA 病毒分类的重要标准。这种新规定已经作为一个常规而被普遍采用。不过要注意国内外仍然有少数学者将转录的模板链称为编码链或有义链这一情况。

人们对于原核生物的转录过程研究得较早，了解得也较多。因此，我们首先讲解原核生物的转录过程。然后再讲解真核生物的转录方面的有关特征。

## 第二节 RNA 合成的酶学[5~14]

### 一、RNA 合成的基本特征

1. RNA 的前体是四种核糖核苷三磷酸(rNTP)：ATP、GTP、CTP 和 UTP。在每一个 NTP 上，有二个 -OH，一个在 2′-C 上，一个在 3′-C 上。

2. RNA 链的生长方向也是从 5′→3′，核苷三磷酸加到新生链的 3′ 端，同时除去一分子焦磷酸而生成磷酸二酯键。焦磷酸又分解为无机磷酸，从而推动反应向聚合方向转移。这与 DNA 复制时的情况基本相同。

3. 转录必须以一条 DNA 链为模板，按照碱基互补的原则进行转录，因此 DNA 中的 A、G、C、T 将分别转录成 U、C、G、A。在一个转录区内，一般只有一条 DNA 链可以被转录。正在转录的 DNA 都是双链，因而正在转录的区域的 DNA 的链必须暂时解开。新合成的 RNA 上的少数几个核苷酸与模板互补，随着链的增长和 RNA 聚合酶的向前移动，原来解链部分的 DNA 又恢复双链结构，而 RNA 逐步游离于 DNA 分子之外。

4. RNA 聚合酶与 DNA 聚合酶不同，RNA 聚合酶能起始一条新链的合成，起始的核苷酸一般是嘌呤核苷三磷酸，而且将在 RNA 链的 5′ 末端保持这一三磷酸基团。

RNA 聚合反应可以用下面的反应式表示：

$$n\mathrm{NTP} + \mathrm{XTP} \xrightarrow[\mathrm{DNA,\ Mg^{++}}]{\mathrm{RNA\ 聚合酶}} (\mathrm{NMP})_n - \mathrm{XTP} + n\mathrm{PPi}$$

其中 NTP 为四种核苷三磷酸前体；XTP 为 RNA 链 5′ 末端的核苷酸，即 RNA 链起始的第一个核苷酸；ppi 代表焦磷酸。

在实验室中可以用 TCA(三氯醋酸)沉淀的方法追踪 RNA 的合成。体系中包括 DNA 模板，$Mg^{++}$、RNA 聚合酶，四种核苷三磷酸前体，其中至少一种前体用放射性进行标记 ($^3H$、$^{14}C$ 标记或 $^{32}P$ 标记于 α-磷原子上)。在保温一定时间后加入 TCA，RNA 不溶于 TCA 而生成沉淀，而前体分子不沉淀，用滤纸过滤。在滤纸上的沉淀中的放射性数量与合成的 RNA 数量成正比。

RNA 合成主要包括四个步骤：(1)RNA 聚合酶结合于 DNA 上的特定位点；(2)起始；(3)链的延长；(4)链的终止和释放。要讨论这些步骤，必须先理解 RNA 聚合酶的结构。*E.coli* 的 RNA 聚合酶是人们最了解的 RNA 聚合酶，其他原核生物的 RNA 聚合酶也十分类似，而哺乳动物的 RNA 聚合酶则有很大的区别。

### 二、*E. coli* RNA 聚合酶

*E.coli* RNA 聚合酶由五个亚基所组成，即 $\alpha_2\beta\beta'\sigma$。$\alpha$、$\beta$、$\beta'$、$\sigma$ 四种亚基的分子量分别为

36.5KDa、150KDa、160KDa、82KDa，整个酶分子的分子量为465KDa，是目前已知的最大的酶之一。$\alpha$、$\beta$、$\beta'$、$\sigma$ 分别是基因 *rpoA*、*rpoB*、*rpoC* 和 *rpoD* 的产物。有时还可以看到与 RNA 聚合酶相结合的一个很小的蛋白质(MW = 10KDa)，叫做 $\omega$ 亚基，其功能尚不清楚，也有人认为，$\omega$ 亚基对于 RNA 聚合酶的结构和功能没有太大的影响。后来，人们还发现 *E.coli* RNA 聚合酶的另一个重要的亚基 NusA，这留待第五节叙述。

在 RNA 聚合的一定阶段。$\sigma$ 亚基将从整个的酶分子上解离下来。在实验室中，用磷酸纤维素柱也容易将 $\sigma$ 亚基与其余部分分开。$\alpha_2\beta\beta'\sigma$ 这样的整个酶分子称为"全酶"，而 $\sigma$ 亚基解离后的其余部分($\alpha_2\beta\beta'$)称为核心酶。当我们讲 RNA 聚合酶时即意味着"全酶"。$\sigma$ 亚基的最主要的功能是识别启动子，特别是启动子上的 R 位点。细胞内哪条链被转录(即转录的方向)，转录的起点的选择都与 $\sigma$ 亚基有关。体外转录实验中，用全酶可得到与体内相同的结果；而用核心酶时，模板链和起始点的选择都有很大的随意性。事实上，核心酶与 DNA 有一定的亲和力，这种亲和力主要来自蛋白质中碱性基团与 DNA 的磷酸根之间的静电引力，是一种与 DNA 序列无关的非特异性结合。这种结合又称为松弛结合。松弛复合物的结合常数为 $2\times10^{11}M^{-1}$，半衰期约为 60 分钟。$\sigma$ 亚基与核心酶结合，则使核心酶产生构象上的变化。全酶与非特异性 DNA 序列的结合力下降上万倍，结合常数约为 $10^7M^{-1}$，半衰期在一秒钟以下；而全酶对于启动子的结合力大大提高，在牢固结合位点上的结合常数高达 $10^{14}M^{-1}$ 数量级，半衰期可长达几个小时。

在核心酶中的四个亚基，其各自功能还不完全清楚。在 *E.coli* 细胞内，某些药物能有效地抑制基因转录，如利福霉素类药物和链霉溶菌素(streptolydigin)。前者能抑制 RNA 合成的起始，后者能抑制 RNA 链的延伸。实验也表明，这两种药物均作用于 $\beta$ 亚基。同时还发现 $\beta$ 亚基与 RNA 的前体核苷三磷酸具有很高的亲和力。这些都表明 $\beta$ 亚基可能参与底物（包括前体核苷三磷酸以及已经形成的 RNA 链)的结合以及催化磷酸二酯键的形成。在离体转录实验中，肝素(heparin)能抑制转录作用，人们发现肝素是与 $\beta'$ 亚基紧密结合的。$\beta'$ 亚基是碱性最强的亚基，而肝素是一种酸性的粘多糖，正好与核酸竞争 $\beta'$ 亚基从而妨碍 $\beta'$ 亚基与有义链的结合。游离的 $\alpha$ 亚基以二聚体的形式存在。在大肠杆菌被 T4 噬菌体侵染时，$\alpha$ 亚基上的一个精氨酸发生了 ADP 核糖基化的修饰，从而降低了 RNA 聚合酶(全酶)对于寄主启动子的亲和力，而同时提高了与 T4 基因启动子的亲和力。这就表明，$\alpha$ 亚基可能参与全酶和启动子的牢固结合。这一牢固结合需要 DNA 双螺旋的局部解链；当 RNA 聚合酶核心酶沿着模板移动进行 RNA 链的延伸时，需要不断地在前面解开双螺旋，在后面恢复双螺旋，这些作用可能与两个 $\alpha$ 亚基的功能有关。

在全酶中，$\sigma$ 亚基究竟与哪一个亚基结合现在还不清楚。全酶重建实验(reconstitution)的步骤是：$2\alpha+\beta\rightarrow\alpha_2\beta$，$\alpha_2\beta+\beta'\xrightarrow{\sigma}\alpha_2\beta\beta'$，$\alpha_2\beta\beta'+\sigma\rightarrow\alpha_2\beta\beta'\sigma$。这就表明，可能核心酶的三种亚基都参与了 $\sigma$ 亚基的结合。

RNA 聚合酶仅仅是复杂的转录机构的核心部分。除了 RNA 聚合酶之外，还需要其他一些辅助的蛋白质因子。如在转录终止时发生作用的释放因子($\rho$ 因子)和抗终止因子等。参与起始作用的 $\sigma$ 因子习惯上算作 RNA 聚合酶的成分，其实也是一种辅助因子。不同的基因种类可能有不同的辅助因子，例如 *E.coli* 的一种热震惊蛋白就是一种特殊的 $\sigma$ 亚基，负责热震惊基因启动子的识别。许多噬菌体也会产生某种因子以修饰寄主 RNA 聚合酶。枯草杆菌就是用各种 $\sigma$ 亚基替换的办法来控制不同发育阶段的基因的表达。

### 三、真核生物的 RNA 聚合酶

尽管世界上第一个 RNA 聚合酶是从哺乳动物细胞中分离出来的，但是直到如今，人们关于真核细胞的了解远远不如对 *E.coli* RNA 聚合酶那样清楚。

真核细胞中有三种 RNA 聚合酶，即 RNA 聚合酶 I、RNA 聚合酶 Ⅱ、RNA 聚合酶 Ⅲ。这些名称最早是依据它们从 DEAE-纤维素柱上洗脱的先后顺序而定出来的。后来发现不同的生物的三种 RNA 聚合酶的洗脱顺序并不相同，因而改用三种不同的 RNA 聚合酶对于 α-鹅膏蕈碱(α-amanitine)的敏感性不同来进行区别。RNA 聚合酶 I 基本不受 α-鹅膏蕈碱的抑制，在大于$10^{-3}$ mol/L 时才表现出轻微的抑制作用；RNA 聚合酶 Ⅱ 对于 α-鹅膏蕈碱最为敏感，在 $10^{-9}$～$10^{-8}$ mol/L浓度下就会被抑制；RNA聚合酶Ⅲ的敏感性介于 I、Ⅱ 之间，在$10^{-5}$～$10^{-4}$ 时表现抑制作用。RNA 聚合酶I存在于核仁中，其功能是合成 5.8S rRNA、18S rRNA 和 28S rRNA；RNA 聚合酶 Ⅱ 存在于核质中，其功能是合成 mRNA 以及 snRNA；RNA 聚合酶Ⅲ也存在于核质中，其功能是合成 tRNA 和 5S rRNA 以及转录 Alu 序列。在细胞质中也能发现一些 RNA 聚合酶 Ⅲ，它是从细胞核中渗漏出来的。三种主要的 RNA 聚合酶的分子量都在 500KDa 左右(14S～15S)，每种酶分子含有两个大亚基和 4～8 个小亚基，每个小亚基的分子量为 10KDa～90KDa。各个亚基的功能还没有真正的了解。真核生物也像原核生物一样，不同种类的基因需要不同的蛋白质辅助因子协助 RNA 聚合酶进行工作。例如，非洲爪蟾的RNA 聚合酶Ⅲ在转录 5S RNA 基因时，必需有一种分子量为 37 KDa 的蛋白质因子，而转录其他基因则不需要这一种蛋白质因子。此外在真核生物的线粒体和叶绿体中，均已发现少数 RNA 聚合酶，它们分子量小，活性也较低，这是与细胞器 DNA 的简单性相适应的。这些 RNA 聚合酶都是核基因编码，并在细胞质中合成以后再运送到细胞器中的。

## 第三节　控制转录起始的 DNA 序列——操作子和启动子的结构[12~29]

### 一、操纵元(operon)及其结构

在原核生物中，很多功能上相关的基因前后相连成串，由一个共同的控制区进行转录的控制。包括结构基因和控制区以及调节基因的整个核苷酸序列叫做操纵元。操纵元的控制区由两个部分组成。就乳糖操纵元来说，从结构基因溯流而上而紧靠结构基因的部分叫做操作子(operator)。它是调节基因(i 基因)产物阻遏蛋白的结合位点。阻遏蛋白本身有两个结合位点，一个是与操作子结合的，叫做操作子位点；另一个是与诱导物结合的，叫做诱导物位点。当细胞中没有诱导物时，阻遏蛋白与操作子结合，而使操作子及后面的结构基因处于关闭状态；当细胞中有诱导物存在时(一般来说，诱导物的浓度＞＞操作子的浓度)，则阻遏蛋白与操作子解离并与诱导物结合，这时操作子及后面的结构基因处于开放状态，使 RNA 聚合酶能够通过而执行其转录功能。操作子好比一道电气开关，控制其后面的全部结构基因的开闭。从操作子再溯流而上，就是另一个控制区域，叫做启动子(promoter)。启动子对于转录的起始十分重要，下面将专门讨论启动子结构。*E.coli* 的乳糖操纵元的启动子和操作子的结构示于图 6-1。

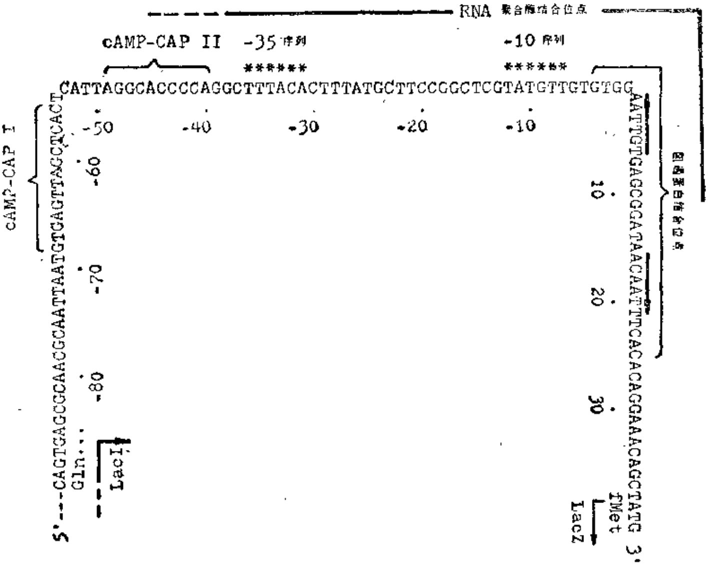



图 6-1 大肠杆菌乳糖操纵元的启动子和操作子的结构及其核苷酸序列(非模板链)

## 二、原核生物的启动子结构

在操纵元中，从 mRNA 开始转录的位点以上，都是启动子的序列。长度从 20bp 到 200 bp 不等。*E.coli* 的乳糖操纵元的启动子有 85 个核苷酸(图 6-1)。整个启动子可以分两个部分，上游部分是 CAP-cAMP 结合位点，下游部分是 RNA 聚合酶进入位点。每个位点又可以分为两部分。CAP-cAMP 结合位点包括了位点Ⅰ和位点Ⅱ。RNA 聚合酶进入位点包括了识别位点和结合位点。

转录过程中有几个重要的事件发生在启动子上，RNA 聚合酶识别 DNA 上的特殊序列，以合适的构象相互结合，打开 DNA 链以进入需要转录的碱基部位，然后开始合成。这一系列事件，主要是由启动子的碱基序列和 RNA 聚合酶的 $\sigma$ 亚基所制导的。对于某些启动子来说，还需要一些辅助性蛋白质。这些事件可以分为三步：(1)RNA 聚合酶结合到识别位点上，(2)移动到一个起始位点上；(3)建立一个所谓开放性启动子复合物(open-promoter complex)。现在人们已经用 DNAase 法和硫酸二甲酯法检定了很多启动子，确定了受 RNA 聚合酶保护的片断和与 RNA 聚合酶紧密接触的位点。现在，人们已经弄清楚数以百计的启动子的核苷酸序列。为了方便，人们将 mRNA 开始转录的一个碱基(转录起始点)定为 +1，沿转录方向顺流而下均用正值表示；溯流而上的启动子部分，均用负值表示(图 6-1)。

1. Pribnow 框　图 6-2 是若干 *E.coli* 和 *E.coli* 噬菌体的启动子的部分核苷酸序列(它们都能为 *E.coli* RNA 聚合酶所识别的)。人们发现在 -10左右的一段核苷酸序列中，大多包含 TATAAT序列或是稍有不同的变化形式，如果用TATPuAT 来表示，则更为适用。其中带

底线的T称为保守T，它存在于目前已知的几乎所有启动子中，一般位于－6到－9。头两个核苷酸是TA的也占3/4以上。这样的六核苷酸序列称为Pribnow框，由于在－10位点附近，所以又称为－10序列。不同的启动子，其位置略有不同，一般都在－4到－13的范围之内。Pribnow框的一致序列为 $T_{80}A_{95}T_{45}A_{60}A_{50}T_{96}$。

| | | Pribnow 框 | mRNA 开始（最后碱基） |
|---|---|---|---|
| lac-UV-5 | CG | TATAATG | TGTGGA |
| fdG2 | AC | TATAATA | GACAGGG |
| $tRNA_1^{Tyr}$ | GA | TATGATG | CGCCCCG |
| $P_R$ | GT | GATAATG | GTTGCA |
| T7A3 | GG | TACGATG | TACCACA |
| lac (w) | CG | TATGTTG | TGTGGA |
| $P_L$ | GT | GATACTG | AGCACA |
| T7A2 | AG | TAAGATA | CAAATCG |
| SV40 | CT | TATAATG | GTTACA |

一致序列： $T_{80}A_{95}T_{45}A_{60}A_{50}T_{96}$

图 6-2　若干启动子的 pribnow 框

Pribnow框是RNA聚合酶的牢固结合位点，简称结合位点。在这里，可能由于RNA聚合酶的诱导作用，在这富含A·T的Pribnow框内的DNA双螺旋首先"溶解"。这个泡状物扩大到17个核苷酸左右，与RNA聚合酶形成所谓开放性启动子复合物。从而使RNA聚合酶定向。使之按顺流方向移动而行使其转录功能。

2. Sextama 框　对于大多数启动子来说，在RNA聚合酶覆盖的部分还有一个重要的区域，叫做Sextama框，其位置在－35附近，因此又叫－35序列（图6-3），其一致序列为：$T_{82}T_{84}G_{78}A_{65}C_{54}A_{45}$。－35序列是RNA聚合酶初始结合位点，RNA聚合酶依靠其 σ 亚基识别该位点，因此又称为RNA聚合酶识别位点。其实验证据是：当RNA聚合酶所保护的片断提纯以后再加入RNA聚合酶，二者仍能很好地结合；但在加RNA聚合酶之前，用Hind Ⅱ在－35序列上切断DNA，则RNA聚合酶就不能结合。同时，目前已经发现的影响启动子活性的若干突变体的突变位点集中在－35序列中和－10序列附近（图6-4）。可见这是两个重要的位点。RNA聚合酶先结合于－35序列，然后才结合于－10序列。至于RNA聚合酶如何从一个位置上转到另一个位置上，过去曾经流行一种"滑动"的假说，现在虽不能排除这种假说，但不太像是滑动。有些实验证据表明，可能是 σ 亚基识别－35序列并与之结合。由于RNA聚合酶分子很长，大约能覆盖70bp的DNA序列。因此酶分子上的一个适合部位就能达到－10

-40 Sextama -20 Pribnow +1

1. lac CCAGGCTTTACACTTTATGCTTCCGGCTCGTATGTTGTGTGGAATTG

2. fd G2 CTTTTTGATGCAATTCGCTTTGCTTCTGACTATAATAGACAGGGTAA

3. $\lambda P_L$ GGCGGTGTTGACATAAATACCACTGGCGGTGATACTGAGCACATCAG

4. $\lambda P_R$ GTGCGTGTTGACTATTTTACCTCTGGCGGTGATAATGGTTGCATGTA

5. $tRNA1^{Tyr}$ CGTAACACTTTACAGCGGCCGCGTCATTTGATATGATGCGCCCCGCTT

$T_{82}T_{84}G_{78}A_{65}C_{54}A_{45}$ .... 17bp .... $T_{80}A_{95}T_{45}A_{60}A_{50}T_{96}$

图 6-3 几种启动子的 Sextama 框，Pribnow 框以及二者之间的距离
图中两条竖线表示 Hind II 切点

-40 -30 -20 -10 +1

GGCACCCCAGGCTTTACACTTTATGCTTCCGGCTCGTATGTTGTGTGGAATTG

△ A A A AA

图 6-4 影响启动子活性的六个 lac 启动子突变体，△表示碱基缺失

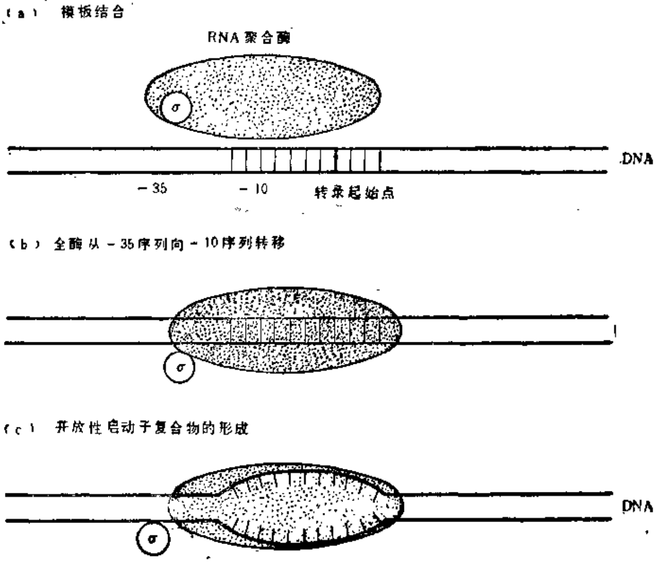



图 6-5 开放性启动子复合物的形成示意图
粗竖线表示 mRNA 开始合成位置，双链熔解范围大约从 -14 到 +2

序列区域。酶分子一旦与－10序列结合以后，就立即从识别位点上解离下来(图 6-5)。

－35 序列的重要性还在于，这一序列的核苷酸结构在很大程度上决定了启动子的强度。RNA 聚合酶很容易识别强启动子，而对弱启动子的识别较差。而 Pridnow 框的碱基序列通过影响开放性启动子复合物的形成速度而控制转录。这两个序列是决定启动子强度的重要因素。这样，细胞可以由此来调节单位时间内所转录的 mRNA 分子数，从而控制蛋白质的合成速度。

在 Pribnow 框和 Sextama 框之间的碱基序列并不特别重要，因为这个序列上的碱基替代突变对转录效率影响不大。然而这两个序列之间的距离却十分重要。实验表明，两个序列之间为 17bp 时转录效率最高。天然启动子中这一段距离大多为 15～20bp，距离的大小也可能是决定启动子强度的因素之一。

3．CAP 位点　上面我们已经讲过，调节基因所产生的阻遏蛋白通过与操作子以及诱导物(例如乳糖操纵元中的别乳糖)相互作用而实现对基因转录的调控。这是一种负调控机制。细胞中许多基因的表达还存在着一种正调控机制。例如，如果*E.coli*在含有乳糖和葡萄糖的培养基中生长，则细菌只利用葡萄糖而对乳糖的存在不予理睬。原来是细胞内缺少葡萄糖时，腺苷酸环化酶将ATP转变成 cAMP，cAMP与其受体蛋白CAP(cyclic AMP receptor protein)结合成复合物，这个复合物再与启动子上的CAP位点结合。这样，启动子上的进入位点方能与 RNA 聚合酶相结合。相反，当有葡萄糖存在时，cAMP 不能形成，CAP 也就不能与启动子上的 CAP 位点相结合，启动子上的 RNA 聚合酶进入位点就不能结合 RNA 聚合酶。这样，一个操纵元中的一组基因就有两道控制开关，只有两道开关同时打开时，基因才能够转录。

乳糖启动子有两个 CAP 结合位点，一个是在－70 到－50(位点Ⅰ)，另一个是在－50 到－40(位 点Ⅱ)。位点Ⅰ包含一个反向重复序列(图 6-6)，这似乎是大多数强结合位点的特征。位点 Ⅱ 是一个很弱的结合位点，但是当 cAMP-CAP 复合结合于位点Ⅰ时，位点 Ⅱ 结合 cAMP-CAP 复合物的能力显著提高。这也就是所谓协同效应(cooperativity)。一旦位点Ⅱ被占据，RNA 聚合酶就很快与－35 序列结合，然后再与－10 序列结合(图6-7)。

AATGTGAGTT | AGCTCACTCA
TTACACTCAA | TCGAGTGAGT

图 6-6　乳糖启动子的 CAP 位点Ⅰ的反向重复序列，竖线为对称轴

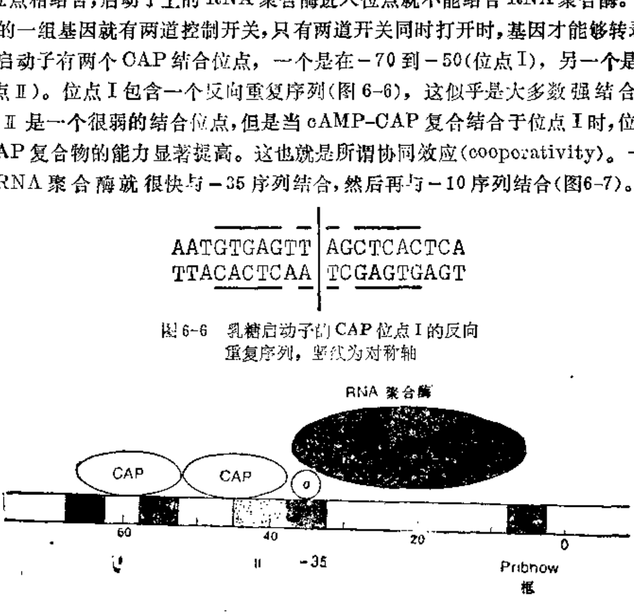



图 6-7　两个 CAP 分子和 RNA 聚合酶分子在乳糖启动子上的排列，Ⅰ和 Ⅱ表示两个CAP 位点

实验表明，如果缺少CAP，尽管乳糖启动子的－35序列和－10序列存在并保持完整，但是－35序列不能很好地结合RNA聚合酶；而Pribnow框确实能结合RNA聚合酶，但是得到的是一个闭合的启动子复合物。也就是说，在这个起始区域内的DNA双螺旋结构不发生熔解作用，开放性启动子复合物不能形成，也就无法开始转录。据认为，乳糖启动子的Pribnow框中的TATGTT序列，其中心的一个G·C碱基对增加了双螺旋结构的稳定性，因此RNA聚合酶还不足以诱导其熔解。CAP-cAMP复合物的结合，使CAP位点Ⅱ附近的富含G·C区域双螺旋结构稳定性降低，因而使Pribnow框的熔解温度降低，从而促进了开放性启动子复合物的形成。关于乳糖操纵元的启动子突变体*lac-UV*-5的研究支持上面的观点。在*lac-UV*-5的Pribnow框中，有两个碱基突变(G→A、T→A)，其序列为TATAAT，在缺乏CAP的情况下，也能形成开放性启动子复合物，进行有效的转录。这可能就是因为中间的G·C碱基对的丧失降低了双螺旋的稳定性，使RNA聚合酶在缺乏CAP的情况下也能诱导其熔解。

使启动子转录效率上升的突变称为启动子上升突变(Promoter up mutation)，使转录效率下降的突变则称为启动子下降突变(Promoter down mutation)。上升与下降是对于同一个启动子在突变前后而言。在不同的启动子中，很可能甲启动子下降突变后的转录效率仍然高于乙启动子上升突变后的转录效率。一般说来，各特异性序列趋向于一致序列的突变以及Pribnow框与Sextama框之间的距离趋向于17bp的突变均表现为上升突变；相反，则产生下降突变。然而，在我们预言启动子突变效应时，也常会失败。很可能因为某种序列特征在那个特定的启动子中并不适用甚至必须排除；也可能在上述几种特异性序列之外的某个碱基或几个碱基对那个启动子是必需的。这些基因的转录很可能还需要有蛋白质辅助因子的参加。

### 三、鉴别RNA聚合酶等蛋白质在DNA上结合位点的方法

RNA聚合酶分子很大，能与DNA分子上许多碱基同时结合。鉴定RNA聚合酶在DNA上的结合位点，对于研究转录过程是十分重要的。

1. DNAase法　RNA聚合酶结合于DNA，然后加入DNA内切核酸酶。核酸酶将大部分DNA都降解成单核苷酸或二核苷酸，而被RNA聚合酶所覆盖的部分则保留下来。然后终止DNAase的作用，除去RNA聚合酶，即可分离到被保护的DNA片断。这种片断一般为40bp左右。但是这种方法很难得出RNA聚合酶所覆盖的DNA分子的真正大小，因为边缘部分的核苷酸对于DNAase常有不同的敏感性。在过量DNAase的作用下，这一部分DNA都被水解掉了。要得到更好的结果，可以采用脚印法(footprinting)，脚印法的大致原理如下，将结合有RNA聚合酶的DNA用限制酶切割，得到一个共同的端点，再将端点的一条链用放射性标记(例如用Klenow片断催化3′端的标记，或用碱性磷酸酯酶催化的5′端的标记)。用微量的DNAaseⅠ部分水解这个酶-DNA复合物。在不受RNA聚合酶覆盖的区域，每个磷酸二酯键都有可能遭到DNAaseⅠ的进攻，但同一个分子上只有少量(如1～2个)切点，将所得到的片断经过碱变性，电泳和放射自显影，就可以知道RNA聚合酶在标记链上覆盖的大小，并可以确定其DNA序列。这个方法的原理与DNA序列分析的原理基本上是相同的。用脚印法得出的RNA聚合酶保护的DNA范围是从－50～－40到＋20这样大约60～70bp。实验还发现，RNA聚合酶与DNA双螺旋的两条链的结合并不对称，这与单链转录是一致的。

用 DNAase 法不能给出 RNA 聚合酶与 DNA 的空间作用位点。而下述的硫酸二甲酯法可以达到这一要求。

2. 硫酸二甲酯法　硫酸二甲酯能使嘌呤甲基化(G 的 7 位甲基化和 A 的 3 位甲基化)，而与 RNA 聚合酶紧密接触的嘌呤就不能被甲基化。当 RNA 聚合酶与 DNA 的一段序列结合时，其中许多碱基并不是处于紧密结合状态，因为两种分子在形状上并不完全互补。在用 DNAase 测得的 RNA 聚合酶保护区内可以找到不被甲基化的嘌呤碱基，这就是与 RNA 聚合酶紧密接触的位点，还可找到可以甲基化的嘌呤碱基，这些核苷酸虽然处于 RNA 聚合酶的覆盖范围内，但并不与 RNA 聚合酶紧密接触。同样通过脚印法和序列分析可以确定这些接触位点。为什么硫酸二甲酯能够区别两种不同的结合状态而 DNAase 却不能？这是由两种分子的相对大小所决定的。因为在 RNA 覆盖区域内，有些部分虽不完全接触，但两种分子之间空隙只有几个 Å。这样的空隙能允许硫酸二甲酯这样的小分子进入，而大分子 DNAase 要接触这些地方的磷酸二酯键却是不可能的。

以上两种方法示于图 6-8。

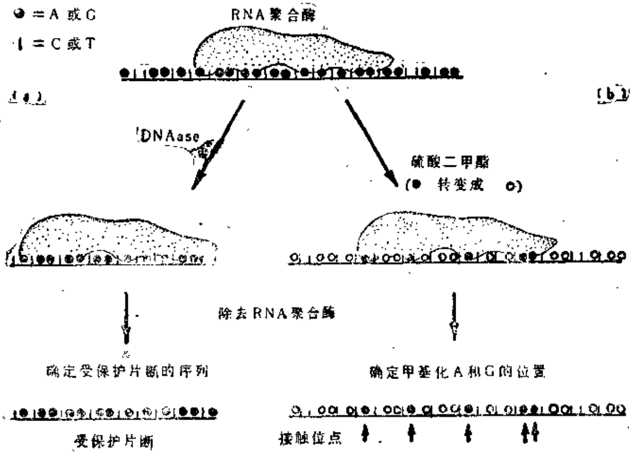



图 6-8　RNA 聚合酶与 DNA 的相互作用。为了清晰起见，图中只画出 DNA 的一条链，实际上是双链 DNA。
(a) DNAase法；(b) 硫酸二甲酯法。

### 四、真核生物的启动子

真核生物有三种 RNA 聚合酶，每一种都有自已的启动子类型。RNA 聚合酶 I 只转录rRNA，只有一种启动子类型。RNA 聚合酶 II 负责蛋白质基因和部分 snRNA 基因的转录，其启动子结构最为复杂。RNA 聚合酶 III 负责转录 tDNA 和 5S rDNA，其启动子位于转录的 DNA 序列之内，称为下游启动子。我们首先讲解 RNA 聚合酶 II 的启动子结构，然后再叙述 RNA 聚合酶 III 的下游启动子。关于 RNA 聚合酶 I 的启动子，将在真核生物基因表达的调控(第九章)中讲授。

1 RNA 聚合酶 Ⅱ 的启动子结构　人们比较了上百个真核生物RNA 聚合酶 Ⅱ 的启动子核苷酸序列，发现真核生物的启动子是多部位结构(multipartite)。主要有四个部位，其中前三个部位是大多数启动子都具备的。

(1) 帽子位点(cap site)，即转录起始位点。其碱基大多为 A(注意：指的是非模板链!)，两侧各有若干个嘧啶核苷酸。这与原核生物 *E.coli* 转录起始点的所谓"CAT 规律"（占相当大的比例，并非全部)相一致。其中 A 为转录起点。

(2) TATA 框：又称 Hogness 框或 Goldberg-Hogness 框。其一致序列为：

$T_{82}A_{97}T_{93}A_{85}\genfrac{}{}{0pt}{}{A_{63}}{T_{37}}A_{83}\genfrac{}{}{0pt}{}{A_{50}}{T_{37}}$，基本上都由 A·T 碱基对所组成，只在很少的启动子中有 G·C 碱基对的存在。然而，在 TATA 框的两侧却倾向于富含 G·C 碱基对的序列，这也是 TATA 框发挥作用的重要因素之一。TATA 框一般位于 -25 附近。除这一点外，其结构与功能均类似于原核生物的 Pribnow 框。离体转录试验表明，TATA 框决定了转录的起始点的选择。也就是说 RNA 聚合酶与 TATA 框牢固结合之后才能开始转录。由于 RNA 聚合酶有相对固定的拓扑形状，其结合位点与催化位点之间的距离也是固定的，这就决定了起始点的选择。体外转录试验表明，如使起始点序列缺失，聚合酶会在同样的距离选择起始点。只不过为了选择 A可以前后偏离 1～2 个碱基而已。天然缺少 TATA 框的基因可以从一个以上的位点开始转录。有些真核基因的启动子没有 TATA 框，例如海胆 H2A 基因。因此有人认为 TATA 框并非转录所必需。然而，绝大多数真核基因都有 TATA 框，而且框内任何一个碱基的替代突变都是强烈的下降突变。因此，这样说可能更妥当些：TATA 框是绝大多数真核基因的正确表达所必需的；在没有 TATA 框的基因表达过程中可能存在着某种替代机制，这种机制可能涉及到某种蛋白质因子与某些 DNA 序列的共同作用。近来发现，鼠类末端脱氧核苷酸转移酶 TdT 基因虽然缺少 TATA 框，但它含有一个 17bp 的称为起始子的序列(initiator，简写作 Inr)，该序列位于转录起始点附近，从 -6 到 +11(GCCCTCATTCTGGAGAC)。实验表明，起始子序列本身能决定转录起始的精确性，但转录效率低；而 TATA 框，SV40 的 GC 框置于起始子序列的上游均能提高转录效率但不改变转录起始的精确性 (S.T.Smale and D. Baltimore, Cell 57: 103, 1989)。

(3) CAAT 框，其一致序列为 $GG^{C}_{T}CAATCT$。一般位于 -75 附近，虽然名为 CAAT 框，但其中头两个 G 的重要性并不亚于 CAAT 部分。如果缺失这两个 GG，则兔子的 β-珠蛋白基因的 CAAT 框变为 TTCCAATCT，其转录效率只有原来的 12%，该框中其他碱基的缺失也导致转录效率的急剧降低。这就表明，CAAT 框可能控制着转录起始的频率。CAAT 框的存在较为普遍。除 CAAT 框之外，真核生物的启动子还有许多其他类型的上游启动子成分，它们可能同样控制着转录起始的频率。

对真核生物启动子的分析要比原核启动子困难得多。真核生物的突变体的诱导和鉴定很难，迄今尚未发现真核启动子的突变体。真核 RNA 聚合酶所保护 DNA 的大小和序列也还没有鉴定出来。目前使用的主要手段是离体转录。

(4) 增强子(enhancer)，又称为远上游序列(far upstream sequence)，一般都在 -100 以上。目前研究得较多的增强子序列主要是SV40的两个正向重复的 72bp 序列(-107～-178和～-179～-250)以及多瘤病毒的 244bp 的增强子序列。SV40 的 72bp 重复序列的增强子作用有如下特点：

a. 对依赖于TATA框的转录和不依赖于TATA框的转录都有增强效应。但对前者的增强效应高。

b. 距离效应：离72bp重复序列近的容易起始转录。

c. 极性效应：转录方向离开72bp重复序列的起始序列优先转录，而朝向72bp重复序列的则效率差。可能是这一重复序列中含有引起转录中止的序列。

d. 细胞类型的选择：在小鼠3T6细胞株中，多瘤病毒的增强子作用高于SV40的增强子；而在HeLa细胞中，SV40增强子则比多瘤病毒增强子作用高5倍。

关于增强子的作用机制。人们一直在进行探索。Chambon等人曾提出，72bp重复序列是RNA聚合酶Ⅱ的双向进入位点。在这里结合之后，RNA聚合酶Ⅱ可以向两个方向滑动，再找到依赖于TATA框的起始点和不依赖于TATA框的起始位点。在离体转录系统中人们并未观察到72bp重复序列的增强作用。奇怪的是，有些基因并不对这一增强子作出反应，如人的α-珠蛋白基因在缺少这一增强子时照样转录，而人的β-珠蛋白基因又强烈地依赖于这种增强子。另一方面，在兔的β-珠蛋白基因下游3 300bp的位置接上72bp重复序列也有增强作用。这是双向进入位点的假说难以解释的。增强子的作用具有细胞或组织的特异性表明，可能在特异性的细胞或组织内存在着增强子的某种调节蛋白质。一个特别的例子是免疫球蛋白基因。在它的转录区域内存在着一个增强子序列，即它是该基因启动子的下游序列增强子。这一增强子只有在B淋巴细胞中才是活跃的，免疫球蛋白也才能表达。另一个惊人的例子是小鼠乳瘤病毒的启动子的上游大约−100bp处，有一个增强子序列，它是固醇类激素与其受体蛋白质所形成的复合物的结合位点。当激素—受体复合物结合到增强子序列上之后，不但能促进其下游方向的基因的转录，而且能促进其上游方向的基因的转录。因此，增强子这一顺式调控成分仍然是通过与之结合的反式调控因子而起作用的(详见第九章)。人们曾经提出Z-DNA与增强子作用机制的关系，但看来可能性不大。

2. RNA聚合酶Ⅲ的下游启动子　在非洲爪蟾5S RNA基因的启动子鉴别以前，人们总是试图在转录起始点的上游方向找出一个基因的启动子。实际上，5S RNA基因的启动子位于转录区内，在转录起始点下游50bp之后。以非洲爪蟾的卵母细胞提取液作为体外转录体系，以不同长度的5S RNA基因(缺失部分以非特异DNA序列代替)为模板进行转录，结果发现，缺失+50以前的序列和缺失+83以后的序列都能正常转录，而唯独+50～+83这段序列不可缺少。把这段DNA序列插入任何DNA中，RNA聚合酶Ⅲ都能识别并起始转录，转录起始点在其上游方向50bp左右的一个嘌呤碱基。这段序列就是5S RNA基因的启动子，这样的位于转录起始点下游的启动子又称为内部启动子。

RNA聚合酶Ⅲ不但能转录5S RNA基因，而且还能转录tRNA基因，部分snRNA基因以及腺病毒的VA基因。这些基因也都是内部启动子。但是，这些基因的启动子与5S RNA基因启动子不同，它们是由不连续的两个区域所组成，靠近5′方向的称为A区，靠近3′方向的称为B区。对于tRNA基因来说，A区相应于tRNA分子的D臂，而B区相应于tRNA分子的TψC臂、在A区和B区之间的序列是可变的，而且有时在反密码环上有内元；但是A区和B区之间的距离有所限制，不能过长。由此可见，各种tRNA结构的相似性，不仅是决定于它们携带氨基酸这一功能的相似性，而且还决定于它们转录的起始和转录后的处理(见后面所述)。有趣的是，原核生物tRNA基因也有大体相同的A区和B区的结构，但这些基因是由上游启动子发起转录的。真核生物的三种RNA聚合酶是何时演化出来的，现在还不清楚。

5S RNA 基因的启动子也有相应的 A 区(+50～+60)和 B 区(+70～+80),但其 B 区的碱基序列与一致序列相差较远。另外，5S RNA 基因的启动子还有另一个重要成分即 C 区(+80～+90)。现将各种 RNA 聚合酶Ⅲ的启动子结构示于图 6-9。

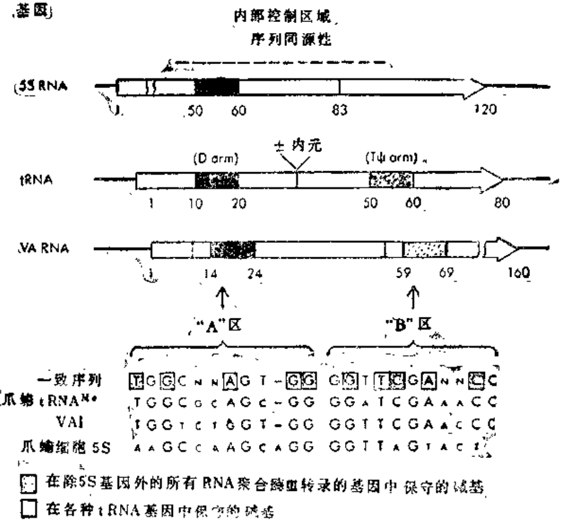



图 6-9　5S RNA,tRNA,VA RNA 基因的启动子结构
及其 A 区和 B 区的碱基序列。

RNA 聚合酶Ⅲ是如何识别不同类别的启动子的呢？上述几类 RNA 的聚合酶Ⅲ的启动子需要共同的转录因子 TFⅢC 和 TFⅢB。对于 tRNA 基因和 VARNA 基因来说，只需要 TFⅢC 和 TFⅢB,甚至有的只需要 TFⅢC 这一个因子。而对于 5S RNA 基因的启动子,除了需要 TFⅢC 和 TFⅢB 之外,还需一种 37KDa 特异的 RNA 聚合酶Ⅲ的转录因子TFⅢA。对于 tRNA 基因来说,TFⅢC主要结合于 B区,其次是 A 区。而对于 5S RNA 基因来说，由于 B 区序列与一致序列相差较多,而必须先有 TFⅢA 与 B 区和 C 区结合，然后 TFⅢC 通过蛋白质与蛋白质的相互作用以及直接与 DNA 相互作用分别与已经结合的 TFⅢA 以及 A 区相结合。在各种Ⅲ类启动子中较保守的 A 区,除了提供转录因子结合的信号外,可能还与 RNA 聚合酶Ⅲ对于转录起点的选择有关，就像 RNA 聚合酶 Ⅱ 启动子中的 TATA 框一样。当 TFⅢC 结合以后，就形成了稳定的预起始复合物Ⅰ,然后 TFⅢB 结合上来就成为预起始复合物 Ⅱ。由于预起始复合物能够稳定地结合于非模板链上，因此能够刺激若干次转录而不是马上被置换。这一过程示于图 6-10。

尽管这几类基因的转录是 RNA 聚合酶Ⅲ、辅助因子和下游启动子共同作用的结果，但是起始位点附近的序列对于转录频率确实有所影响。例如有些 tRNA 基因序列完全相同，但其上游序列不同,造成转录频率的很大差异。

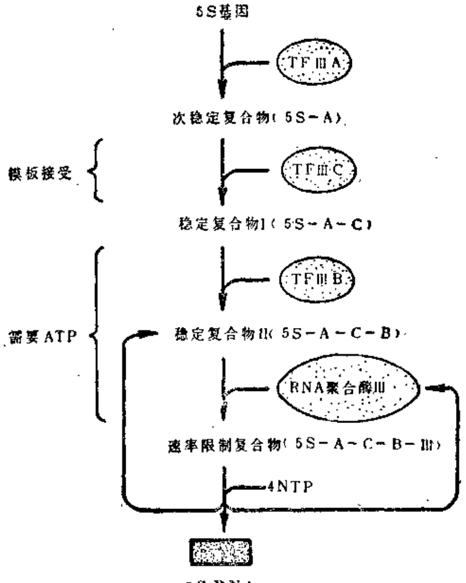



图 6-10　根据离体试验而提出的 5S RNA 基因的启动子与转录因子和 RNA 聚合酶 Ⅲ 相互作用示意图

## 第四节　转录的起始和延伸[11,30~34]

在原核生物中，当 $\sigma$ 亚基与核心酶形成全酶后，全酶与非特异性 DNA 序列的结合力显著下降(结合常数下降)，半衰期极大地缩短。这使得全酶有可能采取快速的尝试——错误的方式(trial and error)去寻找启动子。当 $\sigma$ 亚基发现其识别位点时，全酶就与－35 序列结合(初始结合)，形成一个封闭的启动子复合物。可能由于 RNA 聚合酶全酶的分子很大，其一端可以到达－10 序列，然后由于某种机制，整个酶分子向－10 序列转移并与之牢固结合(此时在操作子上必须没有阻遏蛋白存在)。然后－10 序列及起始位点处发生局部解链，一般为 12～17bp。这时的全酶和启动子的复合物称为开放性启动子复合物。封闭性启动子复合物和开放性启动子复合物均为二元复合物(binary complex)。因为只有全酶和 DNA 这两种成分。在开放性启动子复合物中，RNA 聚合酶上的起始位点和延长位点被相应的核苷酸前体充满，在 $\beta$ 亚基的催化下形成 RNA 的第一个磷酸二酯键。这时由 RNA 聚合酶，DNA 模板和新生的 RNA 链所组成的复合物称为三元复合物(ternary complex)。三元复合物形成之后，$\sigma$ 因子从全酶解离下来，致使三元复合物中核心酶与DNA的亲和力下降到非特异性结合水平以下。这样，既容易在 DNA 链上移动，又因新生 RNA 链参与而造成的三角结合关系（DNA／核心酶－RNA）使核

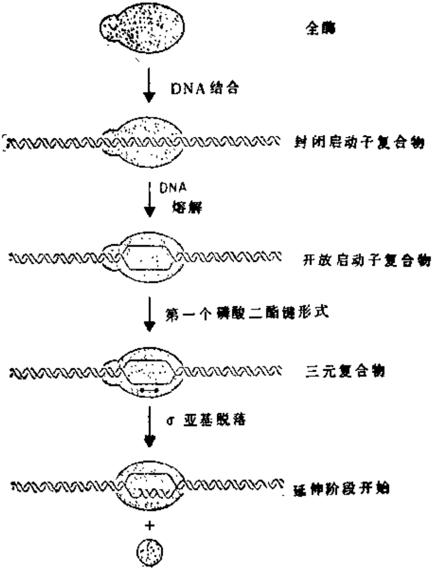


图 6-11 转录的起始过程

心酶继续合成 RNA 链而不致中途脱落。转录起始过程示于图 6-11。

第一个被转录的 DNA 碱基是如何选择的？这个问题还不能确切地回答。这个起始转录的碱基总是在离开保守T大约6～9核苷酸的范围内，其周围的碱基序列并无固定模式，但CAT占相当的比例。据信，这是由于 RNA 聚合酶的结合部位（结合在Pribnow 框上）与聚合活性部位之间的距离大体上是恒定的，因此与保守T的距离比较恒定。模板上的序列只要是 T 或 C 就行了。在 RNA 聚合酶上有两个核苷酸位点，一个是起始核苷酸位点，一个是延长核苷酸位点。只有嘌呤核苷三磷酸充填了起始位点，另外一个核苷三磷酸（可以为任何一种）充填了延长位点并且都与模板互补，才能合成第一个磷酸二酯键。当 RNA 链起始合成之后，起始位点就充当了 RNA 的 3′ 末端位点，而延长位点的职能不变。任何一个磷酸二酯键的形成都必须要使成键的碱基与模板链互补。否则就会被排斥出来。这个功能和催化功能一样，可能都是 $\beta$ 亚基产生的。

利福霉素 SV 可以紧密地结合于 RNA 聚合酶的 $\beta$ 亚基并阻止起始位点的充填。一旦二核苷酸 RNA 形成以后，利福霉素 SV 就不起作用了。而利福平能允许二核苷酸 RNA 的形成。但阻止 RNA 聚合酶的移动；然而，如果在三核苷酸形成以后，利福平也就没有作用了。因此，利福霉素 SV 和利福平都是 RNA 链的起始作用的抑制剂，而不是链的延长的抑制剂。这两种药物广泛地应用于 RNA 合成过程的研究。人们对核心酶的拓扑结构了解得十分有限，只能根据已有资料加以推测，得出一个核心酶与 DNA 和RNA结合关系的可能模式（图6-12）。需要指出，图中任何一个位点都没有在核心酶上明确定位。特别是解旋酶活性（unwindase actvity）和再旋酶活性（rewindase activity）以及有关的拓扑学问题了解得很少。人们的初

步推理是：由于转录复合体携带了新生 RNA 链这条“尾巴”，再加上转录和翻译的耦联，“尾巴”就更加沉重，因此转录复合体不大可能绕着 DNA 分子旋转，而是 DNA 分子本身旋转的可能性较大。DNA 分子旋转带来的超缠问题可能是由 DNA 旋转酶来解决的。人们已经发现 DNA 旋转酶基因的突变妨碍许多基因的正常转录。

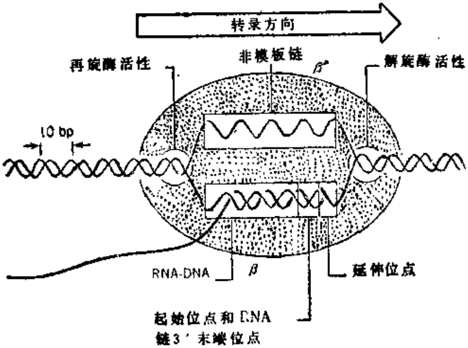



图 6-12　三元复合体的可能的模式（为清晰起见，图中有些部分采用夸张画法，解链的 DNA 多画了三倍，DNA·RNA 杂交双链则多画了 10 倍以上。

新生的 RNA 链与模板形成的杂交双链很短。用胰脏 RNAase 处理转录系统时，由于胰脏 RNAase 能切断单链 RNA 而不能切断 RNA-DNA 杂交分子，结果得到大约 10 个碱基的 RNA 片断。但是这 10 个碱基中多少是受到 DNA 的保护的（形成氢键），多少是受到 RNA 聚合酶保护的（空间作用），现在还不知道。有人推测只有两个左右核苷酸能与 DNA 形成氢键而配对形成杂交状态。而在 DNA 合成中的引物可长达 50～60 个核苷酸均与 DNA 形成氢键结合。为什么有这样的区别，其原因还不清楚。

关于真核生物转录的起始机制，现在还不很清楚。其原因在于转录机构十分复杂，特别是 RNA 聚合酶 Ⅱ 的转录起始。从启动子结构方面来说，它包含了许多启动子成分，每一种成分都有相应的蛋白质因子与之结合，这些蛋白质因子总称为转录因子。在众多的转录因子中，哪种是 Ⅱ 类基因转录所共同需要的？哪些是不同亚类的基因所特有的？在一个基因启动子的许多启动子成分上，是否全要结合有相应的转录因子才能起始转录？这些因子结合的顺序有何影响？是否还有转录起始的抑制因子？这些问题都还没有明确的结论。据信，RNA 聚合酶 Ⅱ 进行的转录除了需要 TATA 框及其蛋白质结合因子TFⅡD外，至少还需要两种启动子成分及其蛋白质结合因子。以上这些问题都已在进行广泛的研究，其起始机制的清晰轮廓不久将会出现在我们面前。

RNA 合成速度大约为每秒 30～50 个核苷酸。但是，RNA 链的延长，并不是以恒定速度进行的，有时会降低速度或延宕。这是延长阶段的重要特点，其原因还不很清楚。人们发现，在通过一个富含 G·C 对的模板以后约 8 至 10 个碱基，则会出现一次延宕。如果在突变体中，G·C→A·T，则减少延宕。如果在连续的G·C对之间只有一个A·T对，把这个 A·T 变为 G·C，则会出现强烈的延宕作用。这种延宕作用可能在 RNA 链的终止和释放过程中起重要作用。

## 第五节 转录的终止[35~52]

转录过程开始以后。RNA 聚合酶就沿着 DNA 移动，继续合成 RNA 链。这条 RNA 链在哪一个位点合成完毕并释放出来呢？这个问题是无法通过比较 RNA 的序列与模板的序列而得出。这是因为细胞中存在着强烈的 RNA 酶活性。3′末端不断遭到降解。在真核生物中，3′末端还要加上 Poly(A)尾巴。poly(A)尾巴并不是直接加在原初转录物的 3′末端，而是 3′末端经过修剪以后才加上去的。因而 RNA 的 3′末端不像 5′末端那样具有明显的标志(三磷酸基团)。目前鉴定转录终止位点都是依靠体外转录体系。由于 RNA 聚合酶的转录作用受到许多参数(如离子强度)的影响，因而体外转录得到的终止点并不能证明体内转录也是在这一点上终止。然而，庆幸的是。人们已经发现了若干种 RNA (原核生物的) 在体外转录中得到的与体内分离到的完全一样。这是人们能够发现终止序列的重要根据。

### 一、终止序列和释放因子

人们通过比较若干原核生物 RNA 终止位点附近的序列。发现终止信号存在于 RNA 聚合酶已经转录过的序列之中。这种提供终止信号的序列就称为终止子(terminator)。终止子可以分为两类。一类是不依赖于蛋白质辅因子而能实现终止作用。另一类是依赖蛋白质辅因子才能实现终止作用。这种蛋白质辅因子称为释放因子，通常又称 ρ 因子。

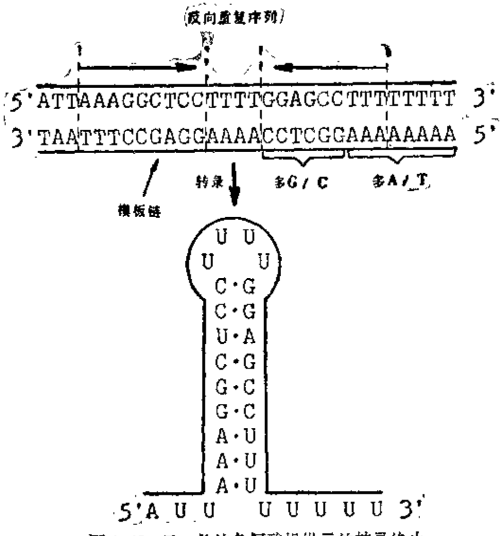



图 6-13 *E.coli* 的色氨酸操纵元的转录终止子序列(不依赖 ρ 因子)

两类终止子有共同的序列特征。在转录终止点之前有一段迴文序列。迴文序列的两个重复部分(每个 7~20bp) 由几个碱基对的不重复节段隔开。迴文序列的对称轴一般距转录终止点 16~24bp。两类终止子的不同点是：不依赖 ρ 因子的终止子的迴文序列中富含G·C碱基对，在迴文序列的下游方向又常有 6~8 个 A·T 碱基对(模板链上为 A) (图 6-13)；而依赖 ρ 因子的终止子中迴文序列的 G·C 对含量较少。在迴文序列下游方向的序列没有固定特征，其 A·T 对含量比前一种终止子低(图 6-14)。

终止子是如何实现终止的作用呢？终止信号是在 DNA 水平上发生作用还是在 RNA 水平上发生作用的呢？上一节中我们讲过，RNA 聚合酶在延长过程合成的 RNA 的速度不是恒定的。在通过一段富含 G·C 对的序列之后约 8~10 碱基则会出现一次延宕。这是了解终止作用机制的重要线索。终止子中的迴

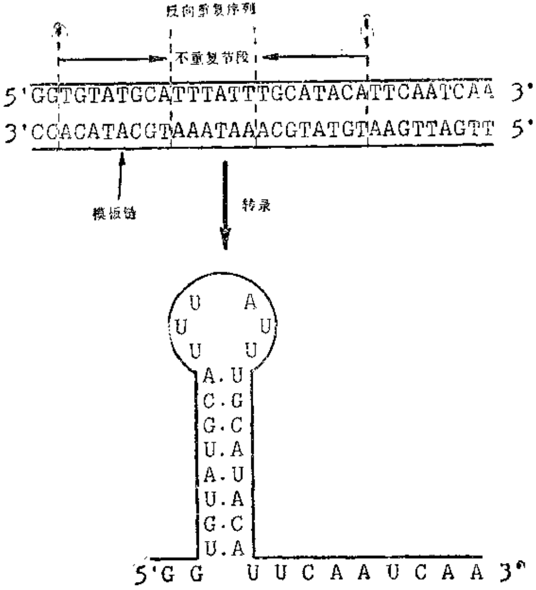



图 6-14 噬菌体 λ 的 $T_{R1}$ 终止子序列(依赖于 ρ 因子)

文序列在 DNA 中形成十字形结构的可能性是很小的，因为从能量上来说是不利的。而且，RNA 聚合酶已经转录了迴文序列，DNA 即使再形成十字形结构对 RNA 聚合酶也不会产生什么影响。可能的应该是在终止子迴文序列处转录出来的 RNA 形成二级结构，即柄-噜噗结构(Stem-loop)，又称发夹结构。这样的柄-噜噗结构可能与 RNA 聚合酶的某种特定的空间结构相嵌合，阻碍了 RNA 链从三元复合体中进一步向外释放，因而造成转录作用的高度延宕。不同的终止子造成延宕的时间长短差异很大。柄-噜噗结构的柄部 G·C 含量越高，延宕时间越长。一个典型的终止子(如色氨酸操纵元的终止子)一般产生 60 秒左右的延宕。单靠柄-噜噗结构，并不足以使转录终止，还需要模板上迴文序列后面的富含 A/U 序列。在色氨酸操纵元的终止子中，由模板上一连串的 A 转录出若干个 U。由于 dA 和 rU 之间的氢键力和碱基堆集力很弱，RNA 和 DNA 杂交部分很容易拆开，于是三元复合物解体，RNA 聚合酶也从 RNA 上解离下来，实现了转录的终止。终止点并不固定在某一个碱基上，有时 RNA 3′ 末端带有 5 个 U，有时 6 个 U。总之，有数个碱基之差(图 6-13)。而依赖于 ρ 因子的终止子不但柄部 G·C 对含量较低(因而延宕时间较短)，而且在迴文序列的下游序列中也常缺乏连续的 A/U。因而，在 RNA 聚合酶转录了迴文序列之后，虽然发生一定时间的延宕，但是如果没有 ρ 因子存在，则 RNA 聚合酶会继续转录下去，这就叫做通读(readthrough)。当 ρ 因子存在时，转录才能终止。需要指出的是，即使是不依赖于 ρ 因子的终止子，在 ρ 因子存在时终止作用更为有效。突变研究表明，减少迴文序列中 G·G 对的含量可以显著地降低终止子的终止效率。用碱基类似物如 ITP 代替 GTP 合成 RNA 时，同样降低终止效率，同时也说明终止子

的终止作用是通过转录成 RNA 序列而实现的。

ρ 因子又是如何导致终止作用的实现的呢？ρ 因子是 *rho* 基因编码的分子量为 55 KDa 的蛋白质，其活性形式为六聚体。在离体条件下。它有两种活性，一种是促进转录终止的活性，另一种是 NTPase 活性。而这后一种活性是实现前一种活性所不可缺少的。ρ 因子是通过识别 DNA 还是 RNA 或 RNA 聚合酶而发挥其作用的呢？实验表明，只有在50碱基以上的RNA链存在时，ρ 因子才有最大的 NTPase 活性。一般认为，ρ 因子可能结合正在合成中的 RNA 链的 5′ 末端，利用水解 NTP 放出的能量从 5′ 向 3′ 方向移动。而 RNA 聚合酶在终止子处的较长时间的延宕给予 ρ 因子以追赶的机会。柄-噜噗结构中柄部的结合力越强，延宕时间也就越长。然后 ρ 因子便与 RNA 聚合酶相互作用而造成转录的终止，RNA 聚合酶和 ρ 因子都从核酸分子上游离出来。然而，并非 RNA 聚合酶在任何地方发生延宕均可造成转录的终止，也就是说，除了 ρ 因子与 RNA 聚合酶相互作用之外，还需要 RNA 与 DNA 的相互作用，也就是说需一定的 DNA 序列。这个序列可能就是紧靠迴文序列下游的一段 DNA 序列。由这段序列转录出来的 RNA 与模板的结合力必须弱到一定的数值，才能配合 ρ 因子与 RNA 聚合酶的作用，实现转录的终止；如果高于这一数值，原则上不能产生终止作用；如果结合力低到某一个更低的数值，则成为不依赖于 ρ 因子的终止子。ρ 因子对于不依赖于 ρ 因子的终止子终止作用的加强就说明了各种终止子由于其序列不同而具有不同强度。可以推想，在细胞内具有同样的 ρ 因子浓度的情况下，各终止子的终止效率也是不同的。因此，细胞中转录终止的实现，不单是产生 RNA 的 3′ 末端，同时也是基因调控的一种途径。ρ 因子的作用和下面要讲的噬菌体 λ 的抗终止蛋白的作用都是对基因表达的调控。

ρ 因子到底如何与 RNA 聚合酶作用？根据突变体的研究，RNA 聚合酶 β 亚基的基因(*rpoB*)的突变可以降低依赖 ρ 因子的终止子的终止作用；RNA 聚合酶 β 亚基基因还有一种突变，能够恢复突变的 ρ 因子的促进转录终止的能力。这些都说明 RNA 聚合酶的 β 亚基参与和 ρ 因子的相互作用。由此可以推测，ρ 因子可能与 RNA 聚合酶的 β 亚基作用，将嵌合于 RNA 聚合酶空间结构中的 RNA 链“抽”出来，也就是 ρ 因子利用水解 NTP 放出的能量，拼命向 3′ 末端移动，与 RNA 聚合酶竞争 RNA 链的 3′ 末端，从而促使 RNA 聚合酶离开 RNA 链和 DNA 链。

ρ 因子是否能与终止子作用，还需看它是否有机会进入 RNA 链的 5′ 末端并沿着 RNA链向3′ 端移动。当翻译作用旺盛时，RNA 5′ 端转录出来，核糖体就会一个接一个地结合上去，ρ 因子很少有插手的机会。即使 ρ 因子结合上去，也只能跟在核糖体后面排队，而 RNA 聚合酶早已通过了依赖于 ρ 因子的终止子。只有当依赖于 ρ 因子的终止子之前发生无义突变，核糖体在无义密码子上解离下来，ρ 因子才有可能追赶上 RNA 聚合酶从而实现终止作用。这也就是为什么同一个转录元里近基因(proximal gene)的一个无义突变能阻止远基因(distal gene)的表达这一极性现象 (polarity) 的道理(图 6-15)。这可能是无义突变的极性效应的主要原因。此外，mRNA 形成的二级结构也可以导致极性现象。*rho* 基因的突变如 *suA*(to suppress polarity)就能够抑制无义突变的极性效应，这是不难解的。当 *suA* 存在时，无义突变就不再造成转录在内部依赖于 ρ 因子的终止子上的终止。

### 二、终止与抗终止及抗终止因子

各种生物对于其不同发育阶段的基因的表达采用不同的控制手段。前面我们已经讲过，

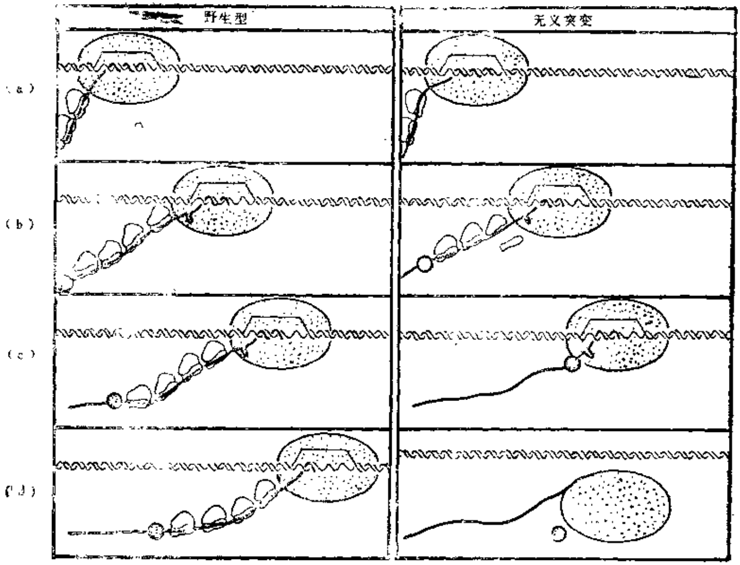



图 6-15　依赖于 $\rho$ 因子的终止子之前发生无义突变，则使突变点以后的基因不能转录，即产生无义突变的极性效应。

(a) RNA 聚合酶后紧跟着核糖体；　(b) $\rho$ 因子与mRNA 结合．在野生型情况下，核糖体跟着 RNA 聚合酶继续前进；而在突变体中，核糖体在无义突变点解离；　(c) RNA 聚合酶在终止子处延宕．在野生型中因核糖体挡路，$\rho$ 因子无法追上 RNA 聚合酶；而突变体中由于核糖体解离而追上 RNA 聚合酶，从而实现转录的终止(实际上是整个操纵元转录的中止)．(d) 野生型中，RNA 聚合酶延宕之后又继续前进，完成整个操纵元的转录，当然翻译作用同样可以完成；而在突变体中，RNA 聚合酶离开了模板，基因的其余部分或后续基因不能得到转录．

枯草杆菌是采用更换 $\sigma$ 亚基的办法；*E.coli* 噬菌体 T7 在基因 1.1 之前的基因均用寄主的 RNA 聚合酶进行转录，而从基因 1.1 开始就使用自己的 RNA 聚合酶（基因 1 产物）进行转录；*E.coli* 噬菌体 λ 则采用依赖于 $\rho$ 因子的终止以及通过抗终止因子的抗终止作用来控制不同阶段的基因的表达。

λ 噬菌体共有四个操纵元：λ 阻遏蛋白操纵元；左向早期操纵元；右向早期操纵元；（右向）晚期操纵元(图6-16)。λ 阻遏蛋白操纵元只含有一个基因即 *cI* 基因，其产物为 λ 阻遏蛋白。左向早期操纵元主要编码与重组有关的蛋白质。右向早期操纵元主要编码与 DNA 复制有关的蛋白质。晚期操纵元编码裂解周期所需的蛋白质，主要为噬菌体头部的结构蛋白，还有溶菌酶等裂解寄主细胞的酶类。晚期操纵元是十分庞大的操纵元，包含 20 多个基因。由于四个操纵元的大小不同。完成操纵元内全部基因转录的时间也不同。在 37℃ 时四个操纵元的全部转录时间分别为 1、6、4、13 分钟(图 6-16)。

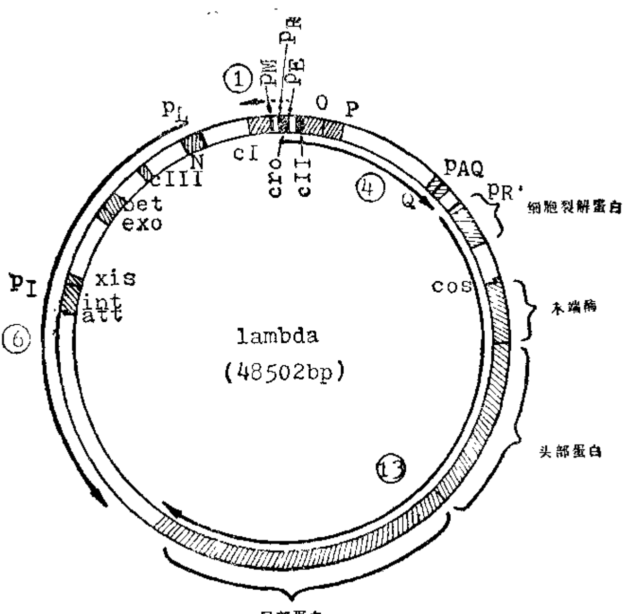



图 6-16　λ噬菌体有四个操纵元

(1) 阻遏蛋白操纵元，启动子为 $P_M$；(2) 左向早期操纵元，启动子为 $P_L$；(3) 右向早期操纵元，启动子为 $P_R$；(4) (右向)晚期操纵元，启动子为 $P_R'$。此外还有三个受到 cII 蛋白调控的启动子 $P_E$,$P_I$ 和 $P_{AQ}$。

在两个早期操纵元内，并非各产生一个 mRNA。左向早期操纵元可以转录出两种mRNA；右向早期操纵元则可以转录出三种 mRNA。这是由于依赖于 ρ 因子的终止作用以及早期基因所编码的抗终止蛋白的抗终止作用而造成的(图 6-17)。在左向早期操纵元中有一个依赖于 ρ 因子的终止子 $t_{L1}$，*E.coli* RNA聚合酶从 $P_L$ 转录至此产生 mRNA L1。在右向早期操纵元中，有两个依赖于 ρ 因子的终止子 $t_{R1}$、$t_{R2}$ 和一个对于抗终止作用不敏感的终止子 $t_{R3}$。寄主 RNA 聚合酶从 $P_R$ 开始转录产生mRNA R1。L1 为极早期基因N的转录产物；R1 为极早期基因 cro(过去称为 tof)的转录产物。两个早期操纵元的其余基因称为迟早期基因。由于极早期基因 N 的表达，其产物蛋白质 N(以下写作 pN)能够阻碍 $t_{L1}$、$t_{R1}$、$t_{R2}$ 的终止作用，才能够使早期操纵元中的迟早期基因得以表达。转录产物分别为 L2、R2、R3。晚期操纵元从 $P_R'$开始转录，在一个依赖于 ρ 因子的 $t_{R4}$ 处终止转录，产生一个 194 个碱基长的6S RNA。6S RNA 不编码任何蛋白质，其功能还不清楚。6S RNA 转录活动很可能是晚期基因转录的前导。由于mRNA R3 中 Q 基因的表达，使转录在 $t_{R4}$ 处不再终止，以使整个晚期基因得以转录产生 mRNA R5(图 6-17)。蛋白质 N 和 Q 即为抗终止因子，又称为抗终止蛋白。

pN 和 pQ 为什么能够抗终止呢？这两个抗终止因子的作用具有高度的特异性。pN 主要作用于 $t_{L1}$ 和 $t_{R1}$ 这两个依赖于 ρ 因子的终止子，然而 pN 对寄主的依赖于 ρ 因子的终止子却

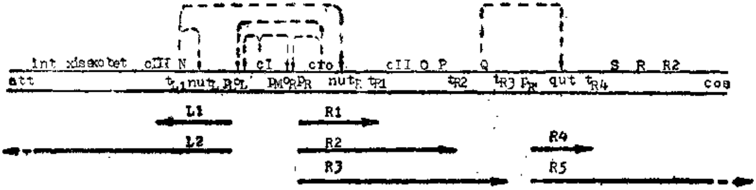


图 6-17 λ噬菌体基因的连瀑式转录。双线代表 λ DNA；基因写在双线上面；DNA 上的作用位点写在双线中间；上面的虚线箭头表示调节蛋白质作用位点；下面的 L1，L2，R1，R2，R3，R4，R5 为 mRNA。

不起作用。如果用寄主的依赖于 ρ 因子的终止子取代 λ 噬菌体的 $t_{L1}$ 或 $t_{R1}$，pN 则又能产生抗终止作用。这说明 pN 的抗终止信号不在终止子上，而是在 λ 终止子的上游某段序列上。现已证明，pN 有两个抗终止信号，一个是 *nutL*，紧接着 $P_L$；另一个是 *nutR*，紧靠 $t_{R1}$。其中 *nut* 即为蛋白质 N 利用位点的意思（pN utilization）。也就是说，两 *nut* 位点对于 pN 所控制转录的基因(N和 *cro*) 的相对位置是极不相同的。*nutL* 位于 N 基因开头，而 *nutR* 则位于 *cro* 基因末尾。两个 *nut* 位点均为 17bp 的序列。左右两个 *nut* 的序列只相差一个碱基对，每个序列中都包含 5bp 的反向重复序列：

AGCCCTGAAPuAAGGGCA

TCGGGACTTPyTTCCCGT

pN 生成以后，首先结合于 *nut* 位点。当 RNA 聚合酶转录通过 *nut* 位点时，pN 便与 RNA 聚合酶结合而修饰了 RNA 聚合酶的构象。使之不再理睬终止子 $t_{L1}$ 或 $t_{R1}$(图 6-18)。

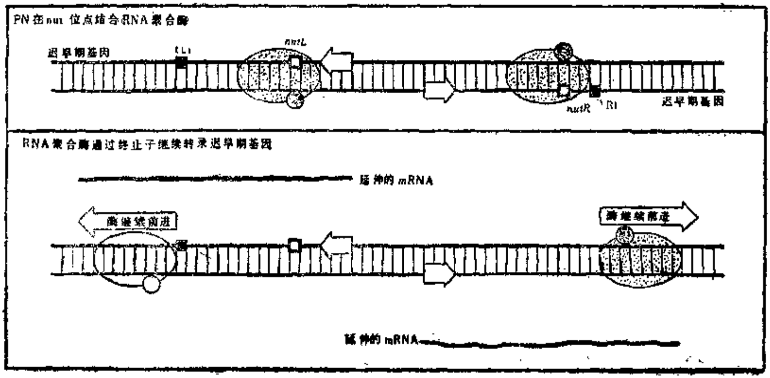


图 6-18 当 RNA 聚合酶通过 *nut* 位点时，pN 与之结合，使酶能够通过终止子 $t_{L1}$ 和 $t_{R1}$ 而转录出包括极早期基因和迟早期基因在内的多基因 mRNA

为了进一步分析 pN 的作用机制，从大肠杆菌中分离到许多可以抑制 pN 抗终止功能的突变体。λ噬菌体不能有效地感染这些突变体，因为除了极早期基因外，λ噬菌体的绝大部分基因不能得到表达。这些突变位点中有一些位于 *rpoB* 基因上，这说明 pN 像 ρ 因子一样，是与

核心酶的 $\beta$ 亚基发生相互作用的。

另一些突变位点位于 *nus*(N utilization substance)位点，其中包括至少三个基因*nusA*，*nusB* 和 *nusE*。后来知道，*nusE* 就是核糖体小亚基蛋白 S10 的基因。

NusA 的功能与 pN 的作用密切相关。λ 噬菌体 N 基因的一些突变体能够克服 *nusA* 突变体对于抗终止的抑制作用。体外实验也表明，NusA 和 pN 很容易相互结合，因为 pN 为分子量 13.5KDa 的强碱性蛋白质，而 NusA 为 69KDa 的酸性蛋白质。NusA 在 $\rho$ 因子存在时，能在许多终止子处有效地进行终止作用。NusA 的作用与 $\rho$ 并不相同。但二者的作用是互补的。NusA 可以与核心酶结合形成 $\alpha_2\beta\beta'$NusA 复合体，但不能与全酶结合。当把 $\sigma$ 亚基加到 $\alpha_2\beta\beta'$NusA 复合体中时，NusA 即解离下来。很可能当全酶合成第一磷酸二酯键后，$\sigma$ 亚基解离，然后 NusA 加入进来形成 $\alpha_2\beta\beta'$NusA 复合体，由这个复合体在 DNA 上移动，NusA 不会解离下来。当转录终止后，这个复合体从 DNA 上解离下来。NusA 或自动解离，或被 $\sigma$ 亚基置换。到底在延长阶段的哪一步上 NusA 加入进来，还不清楚；很可能在延长阶段的任何时刻都可以。这样，NusA 像 $\sigma$ 亚基一样，是作为 RNA 聚合酶的一个亚基成分，一个负责转录起始，一个负责转录终止。

根据现有资料表明，上述 NusA 与核心酶的结合是非序列特异性的，其功能在于促进转录的终止。当这样的复合体遇到 *nut* 位点时，一个分子的 S10，一个分子 NusB，N 蛋白和另一个 NusA 的复合物先后加入进来，与原有的一个 NusA 分子构成转录的延伸控制颗粒 ECP (elongation control particle)。ECP 通过 NusA 分子与聚合酶核心酶相连结，同时还与已经转录出来的 RNA 相连结。ECP 的作用机理还不完全清楚。

### 三、真核生物转录的终止

对于真核生物转录的终止信号和终止机制了解得很少，其主要困难在于很难确定原初转录物的 3′ 末端，因为大多数基因在转录后很快进行加工，无论是 mRNA、tRNA 还是 rRNA 都是如此。就拿 mRNA 来说，转录后 3′ 末端一般都接上 poly(A) 序列。然而，poly(A)并不是直接加到原初转录物的末端，而是要删除一小段，然后才将一个个的腺嘌呤核苷酸聚合上去。我们虽然已经了解产生这个反应的 RNA 末端腺嘌呤核苷酸转移酶 (RNA terminal riboadenylate transferase) 所需的 RNA 序列特征 (AAUAAA)，但仍然无法确切地知道 mRNA 转录的终止点。病毒 SV40 的终止位点已经经过研究，很像 *E.coli* 的不依赖 $\rho$ 因子的终止子，有一个发夹结构，末端带有一连串的 U。人们将不需要 poly(A)尾巴的组蛋白基因突变后注射到爪蟾卵母细胞中。如果突变碱基破坏了发夹环柄部双螺旋结构的形成，则妨碍转录的终止；如果增加柄部双链结构的稳定性，则可更有效地终止。可见，二级结构的形成比具体的碱基序列更为重要。这种二级结构对转录的终止作用同样发生在 RNA 水平上。然而，对于绝大多数具有 poly(A) 尾巴的 mRNA 的转录在何处终止，目前仍然不清楚。据推测，可能需要一段富含 A·T 的序列。

对于 RNA 聚合酶 Ⅲ，体外转录与体内转录产物相同，这就表明体内转录确实是在 RNA 末端处终止的。爪蟾 5S RNA 的 3′ 末端为 4 个 U，而这四个 U 的前后均为富含 G·C 的序列，这种分布在富含 G·C 序列中的寡聚 T(4 个以上) 是所有真核生物 RNA 聚合酶 Ⅲ 转录的终止信号。这种序列特征高度保守，从酵母到人都很相似。任何改变这种序列特征的突变都将导致通读，而且当改变这寡聚 T 序列后，还阻碍产物的转录后处理。小鼠 5S RNA 基因转录终

止发生在一排 23 个 A 处，人类 Alu 序列的转录有 80% 终止在连续的 9 个 A 外。这些序列特征表明，RNA 聚合酶 Ⅲ 能够识别 DNA 上的二级结构松驰的区域或者新生的转录产物与模板结合力较弱。除 RNA 聚合酶Ⅲ本身之外，似乎不需要其他蛋白质因子参与转录的终止作用。

## 第六节 转录后处理(一) 转录产物的修饰——真核生物 mRNA 的帽子和尾巴[53~62]

在真核生物中，几乎全部的成熟 mRNA 都具有 5′ 帽子结构，大部分成熟 mRNA 还同时具有 3′ 多聚 A 尾巴。这是真核生物 mRNA 的重要特点。

### 一、帽子

尽管真核生物的 mRNA 的转录也是以嘌呤核苷三磷酸(pppA 或 pppG)领头的，但是人们用 RNAase 处理成熟的 mRNA 时，并没有得到预期的 pppA 或 pppG，而是得到一个以 5′-5′ 三磷酸二酯键相连的二核苷酸，而且 5′ 端的一个核苷酸总是 $N^7$-甲基鸟苷酸。这样，成熟的 mRNA 分子并没有自由的 5′ 端。mRNA 5′ 端的这种结构就称为帽子。不同的真核生物的 mRNA 具有不同的帽子；同一种真核生物也常具有不同的帽子 (6-19)。现在还不清楚，同一种 mRNA 是否只有一种帽子。上述的 $N^7$-甲基鸟嘌呤核苷酸部分称为帽子 0，其符号的 $m^7GpppX$。单细胞的真核生物(如酵母)只具有帽子 0。如果在原来转录物的第一个核苷酸的 2′-O位上产生甲基化，则构成帽子 1(包括帽子 0 部分)，其符号为 $m^7GpppXm$。这是除了单细胞真核生物外的其余真核生物的主要帽子形式。在高等真核生物中，如果第一个核苷酸是 A 的话(大多数如此)，则在 A 的 $N^6$ 位上还有一个甲基。在有些真核生物中，在第二个核苷酸

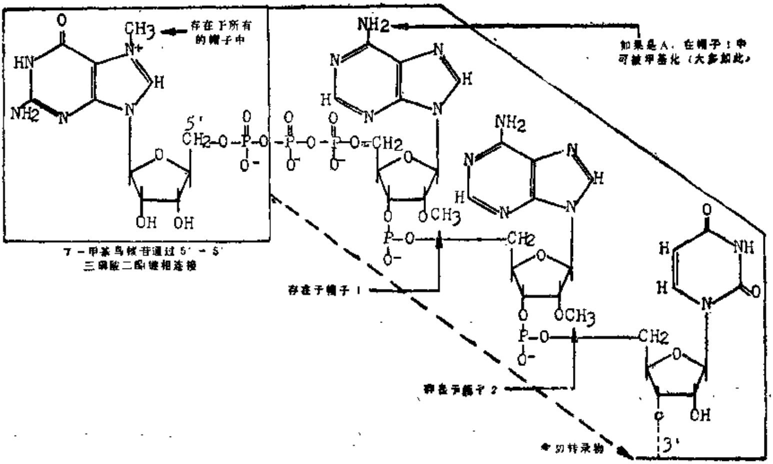



图6-10 真核生物mRNA 5′端的三种不同的帽子结构

(可以为 A,G,C,U中的任何一个)的 2′-O位上还可以再产生甲基化。构成帽子 2(包括帽子 1 部分),其符号为$m^7GpppX_{mp}Y_m$。帽子 2 一般只占戴帽子 mRNA 的 10~15% 以下。

mRNA的5′帽子结构是由一系列的酶促反应生成的,其反应如下:

$$_{ppp}X_pY_pZ \xrightarrow{\text{RNA 三磷酸酯酶}} {}_{pp}X_pY_pZ + P_i \quad (1)$$

$$G_{ppp} + {}_{pp}X_pY_pZ \xrightarrow{\text{RNA 鸟苷酸基转移酶}} G_{ppp}X_pY_pZ + PP_i \quad (2)$$

$$G_{ppp}X_pY_pZ + SAM \xrightarrow{\text{RNA 鸟嘌呤-7-甲基转移酶}} m^7G_{ppp}X_pY_pZ + SAH \quad (3)$$

$$m^7G_{ppp}X_pY_pZ + SAM \xrightarrow{\text{RNA 核苷 2′-0-甲基转移酶}} m^7G_{ppp}X_{mp}Y_pZ + SAH \quad (4)$$

$$m^7G_{ppp}A_{mp}Y_pZ + SAM \xrightarrow{\text{RNA2′-0-甲基腺嘌呤-6-甲基转移酶}} m^7G_{ppp}{}^mA_{mp}Y_pZ + SAH \quad (4')$$

$$m^7G_{ppp}X_{mp}Y_pZ + SAM \xrightarrow{\text{RNA 核苷 2′-0-甲基转移酶}} m^7G_{ppp}X_{mp}Y_{mp}Z + SAH \quad (5)$$

由反应(3)、(4)、(5)分别生成帽子 0、帽子 1 和帽子 2。在这些反应中,甲基供体都为 S-腺苷甲硫氨酸(SAM),失去甲基后变为 S-腺苷高半胱氨酸(SAH)。反应(2)中的 RNA 鸟苷酸基转移酶常被称为戴帽酶(capping enzyme)。

真核生物的 mRNA 为什么需要帽子结构即帽子结构的生物功能是什么?这个问题现在还不完全清楚。根据现有资料,可以归纳为如下几点:(1)帽子结构为核糖体对于 mRNA 的识别提供了信号。帽子 0 部分的结构(包括没有自由 5′ 端的鸟嘌呤核苷酸,N-7 甲基,以及 5′-5′ 三磷酸二酯键)是核糖体识别所必需的,在帽子 1 和帽子 2 结构中的甲基化也能够增进核糖体对于 mRNA 的结合,但并非决定性的。N-7 甲基的作用是引入一个正电荷而不是甲基本身,用乙基代替甲基能同样促进核糖体的结合。体外翻译试验表明,缺少 5′ 帽子的mRNA,不能进行有效的翻译。最近还发现了一种所谓帽子结合蛋白(cap binding protein,简称 CBP)。它在依赖帽子的翻译中起重要作用,脊髓灰质炎病毒的 mRNA 是没有帽子结构的,它的 mRNA 中必然有一种能为蛋白质翻译系统所识别的特征。正因为如此,它的mRNA 的翻译就不需要 CBP,而且通过某种机制干扰 CBP 与寄主 mRNA 的结合,从而抑制寄主的蛋白质合成。(2)帽子结构增加 mRNA 的稳定性,保护 mRNA 免遭 5′ 外切核酸酶的攻击。(3) Klug 等人发现,外加带有帽子的 RNA 能促进流感病毒 RNA (+)链的合成。而且帽子结构和头 12~15 个核苷酸还共价结合于新合成的RNA(+)链的 5′ 端。起这种引发作用的RNA,必需具有帽子 1 或帽子 2 的结构。不但 7-甲基鸟嘌呤及 5′-5′ 三磷酸二酯键是必需的,而且核苷 2′-0-甲基也是必需的。然而这一引发 RNA 复制的功能对于细胞本身有何用处,现在还毫无所知。

此外,在高等真核生物中,除了帽子结构上出现甲基化以外,在 mRNA 结构的内部也以碱基总数 1/1 000 的几率出现甲基化的腺嘌呤。这样,平均长度(1 800~2 000b)的 mRNA 分子可能出现 1~2 个 $N^6$ 甲基化腺嘌呤,其功能还不知道。有些 mRNA,如珠蛋白 mRNA,就完全没有这样的甲基化。

## 二、多聚(A)尾巴

大多数的真核 mRNA 都具有 3′ 末端的多聚(A)尾巴。具有这种特征的 mRNA 表示为 poly(A)$^+$;不具这一特征的则写作 poly(A)$^-$。多聚(A)尾巴大约为 200b。

多聚(A)尾巴不是由 DNA 编码的。而是在一个 300KDa 的 RNA 末端腺苷酸转移酶

(RNA terminal riboadenylate transferase)催化之下，以 ATP 为前体，一个一个地聚合到 mRNA 的 3′ 末端。其反应如下：

$$多聚核糖核苷酸 + nATP \xrightarrow[或Mn^{++}]{Mg^{++}} 多聚核糖核苷酸(A)n + nppi。$$

催化这一反应的 RNA 末端腺苷酸转移酶通常又称为poly(A)聚合酶。其催化的反应是：将 ATP 中的 AMP 残基通过 3′-5′ 磷酸二酯键连接到多聚核糖核苷酸的 3′ 末端。根据这一定义，则排除了诸如末端脱氧核糖核苷酸转移酶，同聚核苷酸聚合酶，依赖引物的核糖核苷酸聚合酶，多聚核苷酸磷酸化酶等不正确的名称。

然而，poly(A)尾巴并不是加在转录终止的 3′ 末端，而是转录产物的 3′ 末端由一个 360KDa 的特异性酶的作用切除一段，该酶能够识别切点上游方向 13～20 碱基附近的序列 AAUAAA 以及切点下游方向的 GUGUGUG (单细胞真核生物除外)。然后再由 poly(A)聚合酶催化加上 poly(A)尾巴(图 6-20)。如果这一识别信号发生突变，如 U→G，则切除作用和多聚腺苷酸化作用均显著降低。是什么机制维持 poly(A)尾巴的长度总是在 100～200b，现在还毫无所知，可能除上述两个酶之外，还需要其他蛋白质成分和 RNA 成分以组成具有完整的多聚腺苷酸化功能的 RNP 颗粒。

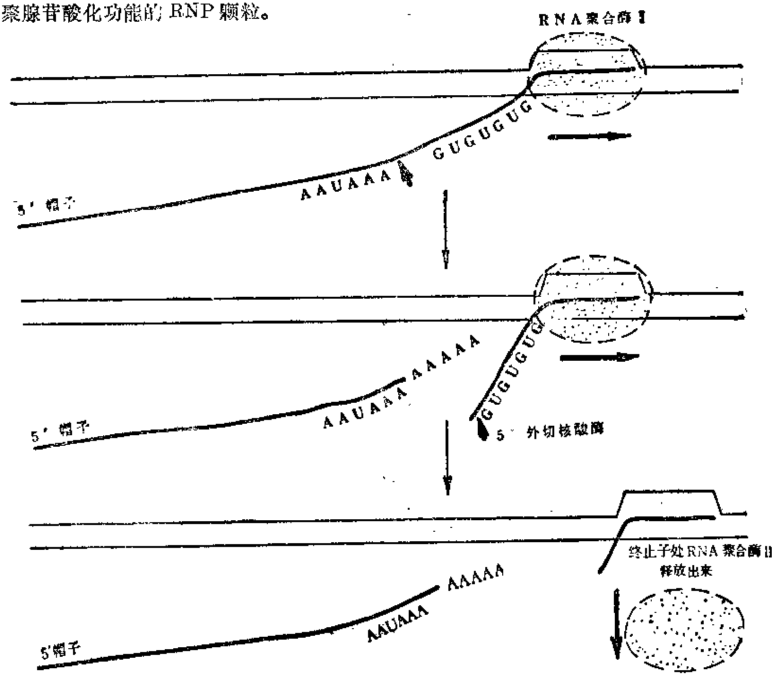



图 6-20 真核生物 RNA 聚合酶 Ⅱ 的转录产物的多聚腺苷酸化与转录的终止。AAUAAA 和 GUGUGUG 为切断 RNA的酶所识别的信号。切断后，mRNA 聚合酶可继续转录直到终止子或者由别的机制促进 RNA 聚合酶Ⅱ的释放。

不但细胞核中有多聚(A)聚合酶，而且细胞质中也有类似的酶。但细胞质中的酶的主要作用是将已有的poly(A)延长，而不是重新起始poly(A)的合成。mRNA的poly(A)尾巴在细胞质中可能受到核酸酶的降解而逐渐变短。这样，由于细胞质中poly(A)聚合酶的作用，可使变短的poly(A)尾巴再变长。

关于poly(A)尾巴的功能问题，尽管人们进行了极其广泛的探索，但现在还是很不清楚。加polp(A)尾巴一般发生RNA拼接之前，但对于拼接作用并无太大的影响。有人推测poly(A)可能与核-质转运有关，但是相当数量(占大约三分之一)的没有poly(A)尾巴的mRNA如组蛋白mRNA，也以照样通过核膜而进入细胞质。在细胞质内，poly(A)的长短对于mRNA的翻译也无明显的影响。在细胞质内，mRNA的poly(A)尾巴与蛋白质相结合，构成mRNP。哺乳动物的这种蛋白质分子量大约为78KDa。mRNP显然不妨碍mRNA的翻译。但是mRNP形式的真正功能是什么还不得而知。

在体外体系中，多聚腺苷酸化可被抗snRNP，特别是抗U1 snRNP的抗体所抑制。这就意味着U1这种核内小RNA分子(small nuclear RNA，简称snRNA)可能与多聚腺苷酸化作用有关。U1可能不是单独起作用，而是可能构成一个更为复杂的复合体而起作用。在海胆的提取液中有一种12S的因子，分子量为250KDa，其中包含一个60碱基长的RNA组分。这一因子能够使海胆组蛋白H3的基因在非洲爪蟾的卵母细胞中转录时得到恰当的3′末端。我们还不知道这种核蛋白是何时起作用以及怎样起作用的，但是使人们更为相信，snRNP可能参与所有的hnRNA的转录后处理。

## 第七节　转录后处理(二)　基因间序列的去除——稳定RNA从多基因转录产物中的分离[63~71]

### 一、不稳定RNA与稳定RNA

在原核生物中，多基因的mRNA生成以后，绝大部分直接作为模板去翻译各个基因所编码的蛋白质，并以此手段协调相关基因的表达量。原核生物的mRNA寿命很短，其半衰期通常为几分钟。这个现象看起来似乎是浪费，而实际上这恰恰是原核生物内的重要调控机制。如果一种酶或蛋白质不再需要，只要简单地关闭其mRNA的合成就行了。在真核生物中，可能有特别的机制来延长需要翻译的mRNA的寿命，而缩短不再需要的mRNA的寿命。而rRNA和tRNA则不同，它们不是翻译的模板，而是蛋白质合成机器的组成成分。不管在原核生物中还是在真核生物中，它们都很稳定，半衰期一般为几个小时。用$^3H$尿嘧啶进行的脉冲追踪试验可以检定各种RNA的寿命(图6-21)。

无论是rRNA还是tRNA，都不是最初转录产物，其证据如下：(1)rRNA和tRNA分子的5′端都是单磷酸，而所有原始转录产物的5′端都是三磷酸；(2) rRNA和tRNA分子都比原始转录产物小，也就是比DNA的转录元(transcripton，亦即转录单元transcription unit)小；(3)所有的tRNA分子都含有原始转录产物所没有的异常碱基，也就是除了A、G、C、U以外的碱基。因此，无论是rRNA，还是tRNA，都需要作转录处理。以删除基因间序列以及基因内的介入序列(内元)。

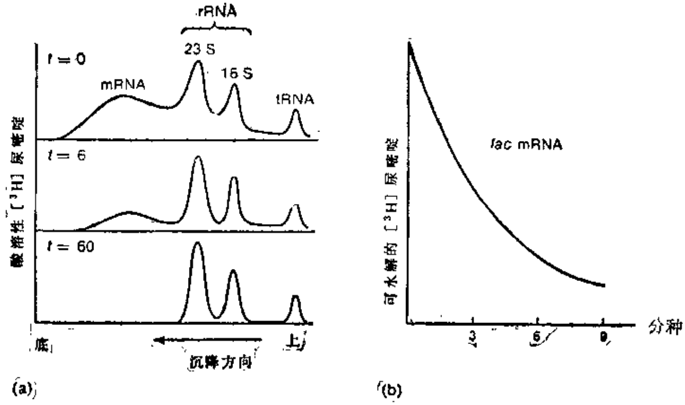



图 6-21 用[$^3$H]尿嘧啶脉冲追踪实验检定 RNA 的寿命

## 二、rRNA 的转录后处理

*E.coli* 共有三种 rRNA，即 5S rRNA，16S rRNA，23S rRNA，它们分别含有 120，1 541，2 904 个核苷酸。rRNA 基因都与 tRNA 基因混杂排在一个操纵元中。*E.coli* 共有七个这样的操纵元。每个操纵元中 tRNA 基因的种类，数量和位置都各不相同，没有固定的模式。

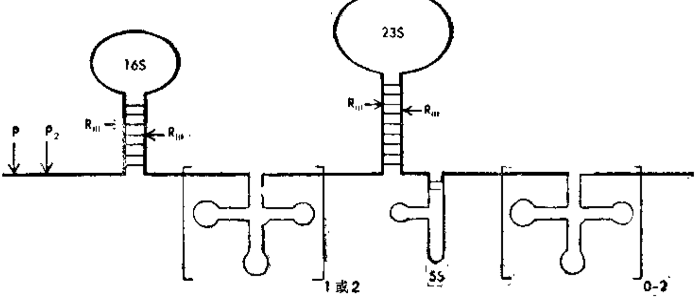



图 6-22 *E.coli* K12 的七个 *rrn* 操纵元。细线部分代表成熟的 rRNA 或 tRNA 的序列，RⅢ 代表 RNAaseⅢ的切点，$P_1$，$P_2$ 代表两个相距约 100 个核苷酸的转录起点。

| 操纵元 | 间 隔 tDNA | 收尾 tDNA | 染色体上位置(分) |
|---|---|---|---|
| *rrnA* | Ile1, Ala1B | — | 86 |
| *rrnB* | Glu2 | — | 89 |
| *rrnC* | Glu2 | Asp1, Trp1 | 84 |
| *rrnD* | Ile1, Ala1B | Thr | 71 |
| *rrnE* | Glu2 | — | 90 |
| *rrnG* | Glu2 | — | 56 |
| *rrnH* | Ile1, Ala1B | Asp1 | 5 |

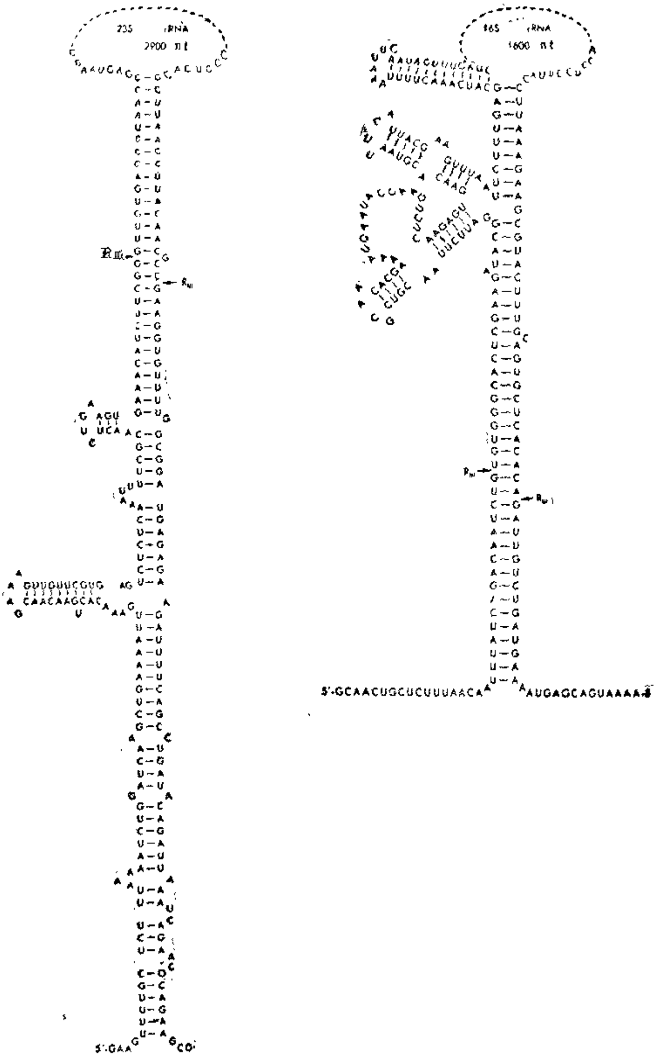



图 6-23 *E. coli* 的 rRNA 基因的原初转录物由 RNAase Ⅲ 切断产生 16S rRNA 的前体 P16 和 23S rRNA 的前体 P23。

但是，三个 rRNA 基因的相对位置都相同，即…16S rDNA，间隔 tDNA，23S rDNA，5S rDNA，收尾 tDNA。*E. coli* K12 株的七个 *rrn* 操纵元的结构示于图 6-22。这七个操纵元大多数集中于复制原点的两侧不远处，而且其转录方向总是与 DNA 复制方向相一致的。

当 *rrn* 操纵元转录产生一个多基因的 30S RNA 后，必须切除先导序列，拖尾序列和基因间序列。将各 rRNA 相互分开。细菌的 RNAaseⅢ执行这一任务。如果 RNAaseⅢ 基因发生突变，则造成 30S RNA的积累。RNAaseⅢ 必须作用于双链 RNA，在两条链上的切点通常错开 2 个碱基对，现在还不知道这两个切点是同时进行的还是有先有后。如果在双链中间有不配对的单链泡，则常常切断单链，而双链仍然是 RNAaseⅢ的识别标志。现在还不知道 RNAaseⅢ要求什么样的序列专一性，可能是识别一级结构和二级结构相结合的某种特征。在 30S RNA 的 16S rRNA 和 23S rRNA 部位的切点都处在双链 RNA 的中间部分，作用后生成比 16S rRNA 和 23S rRNA 略大的前体 P16 和 P23 (图 6-23)。P16 和 P23 再经过适当的加工修剪，就成为 16S rRNA 和 23S rRNA，然而其过程尚不清楚。

真核生物有四种 rRNA，即 5.8S rRNA，18S rRNA，28S rRNA 及5S rRNA，前三者的基因组成一个转录元，产生 45S 的前体 RNA。由于真核生物 rRNA 的加工过程比较缓慢，其中间产物可以从各式各样细胞中分离出来。因此真核生物的 rRNA 加工过程比较清楚。人类的 45S 前体上有 110 多个甲基化位点，几乎全部在核糖性上。这些甲基化的核苷酸序列全部进入成熟的 rRNA，18S rRNA 上有 39 个，28S rRNA 上有 74 个。可见。这些甲基化位点可能是加工过程中酶的识别标志。这种酶很可能类似于原核生物的 RNAaseⅢ，45S RNA 在

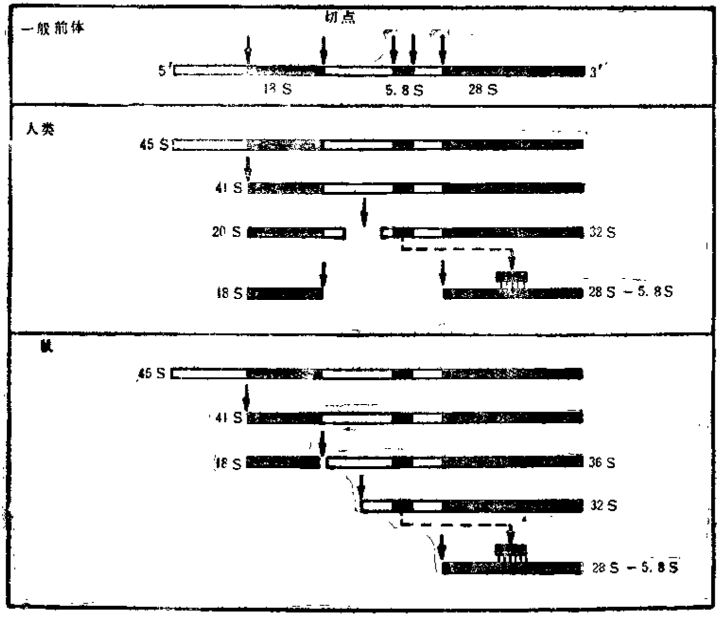



图 6-24 哺乳动物 45S rRNA 前体的转录后处理

转录以后很快与蛋白质结合,因此加工的对象并不是游离的 RNA,而是一种核蛋白。

哺乳动物 45S RNA 的加工有几种不同的方式。例如人类 45S RNA 与小鼠的 45S RNA 的加工方式就不相同(图 6-24)。二者的共同之处是先切除 18S rRNA 之前的 5′ 端序列;然后的一个切点使 18S rRNA 部分与后面的 5.8S 和 28S rRNA 部分分开;最后 5.8S rRNA 和 28S rRNA 通过碱基配对相结合。每个切点的精确位置还不知道,切开以后也需进一步修剪。有些生物 rRNA 的前体含有内元,例如四膜虫 rRNA,其拼接机制在下一节中叙述。

近年来人们发现,哺乳动物的 rRNA 的最原始转录物并不是 45S,而是 47S。由于 3′ 端部分很快进行转录后处理,5′ 也除掉一小部分而变为 45S。还可能存在一种 46S 的中间体,关于真核生物的 rRNA 转录的调控将在第九章中讲述。

### 三、tRNA 的转录后处理

原核生物的 tRNA 基因的转录单元大多数是多基因的。不但同一种 tRNA 的几个基因拷贝可以转录在一条 RNA 中,而且不同的 tRNA 也可以转录在一条 RNA 中。还有的 tRNA 与 rRNA 组成转录单元,上面已经讲过。

从体内合成的角度来说,人们了解得最清楚的 tRNA 分子要算是 *E.coli* 的 $tRNA_1^{Tyr}$ 分子。它是携带酪氨酸的 tRNA,它识别的密码子是 GAU。这个分子含有 85 个核苷酸,它在 *E.coli* 的 DNA 中有两个基因拷贝。每个基因 350 个核苷酸(而不是 85 个)。两个基因之间有一段 200 个核苷酸的间隔子(spacer)。两个基因转录成一条 RNA(图 6-25)。

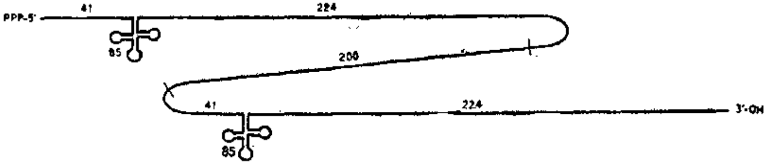



图 6-25 两个 $tRNA_1^{Tyr}$ 基因的原始转录产物

为了简化 $tRNA_1^{Tyr}$ 的合成过程的研究,用遗传学方法产生了只有一个 $tRNA_1^{Tyr}$ 基因的转录单元,也就是在 RNA 分子之前有 41 个核苷酸的前导序列,在 tRNA 分子之后有 224 个核苷酸并有依赖于 $\rho$ 蛋白的终止子。这个原始转录物的修饰过程示于图 6-26。

$tRNA_1^{Tyr}$ 转录后处理过程主要有如下五步:

(1) 一种能识别发夹结构 I 的内切核酸酶在箭头所指处切断。

(2) RNAaseD 能够识别 CCA 末端序列。一个一个地切除七个核苷酸。与这个外切核酸酶作用类似的还有 RNAaseQ。所生成分子叫做前 tRNA,容易从 *E.coli* 中分离到。其结构可以写成 PPP-5′-(41 个核苷酸)-tRNA-(2 个核苷酸)-3′-OH

(3) 一个内切核酸酶 RNAaseP 在箭头所指处切断。据信,RNAaseP 负责 *E.coli* 的所有的 tRNA 原始转录物 5′-P 末端的处理。

(4) RNAaseD 除去 3′ 末端的二个核苷酸,这时的分子已经和成熟的 tRNA 分子具有相同的长度。

(5) 有 6 个碱基通过专一性的酶过程变为异常碱基,其具体反应过程就不在这里叙述了。

众所周知,tRNA 的序列是高度保守的。各种 tRNA 都具有三叶草形结构。tRNA 前

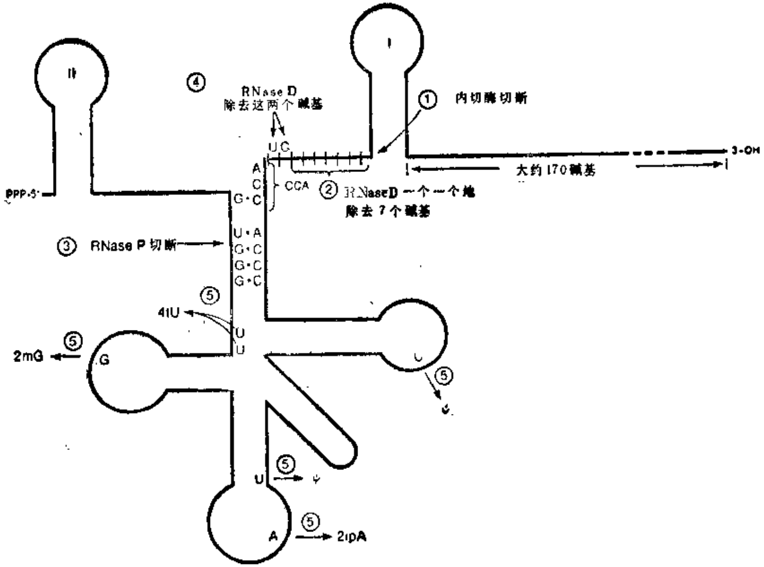



图 6-26 *E.coli* 的 $tRNA_1^{Tyr}$ 基因转录物的转录后处理

ψ=假尿嘧啶核苷

2iPA=2-异戊烯基腺嘌呤核苷

2mG=2-0-甲基鸟嘌呤核苷

4tU=4-硫尿嘧啶核苷

体分子中 tRNA 部分的二级结构，除了反密码子嘌噗外均与成熟的 tRNA 相同。这种保守性不仅反映了它们结合并携带氨基酸这种功能上的共性，而且很重要的方面是反映了它们转录加工方面的共性。真核 tRNA 还反映了转录所需的下游启动子的共性。tRNA 的加工酶系主要包括 tRNA 核苷酸转移酶，RNAaseP，RNAaseD，RNAaseⅢ 以及 $RNAaseP_4$ 等。现将这些酶的主要特性介绍如下。

tRNA 核苷酸转移酶(tRNA nucleotidyl transferase)：该酶能特异地以ATP和 CTP 为前体催化 tRNA3′ 末端的 CCA的生成。tRNA基因有两种类型，一种是有 CCA 序列(I型)，一种是没有 CCA 序列(Ⅱ 型)。但是成熟的有功能的 tRNA 均有 CCA3′-OH 末端，这是 tRNA 结合并携带氨基酸的功能所必需的。原核生物的 tRNA 基因绝大部分为 I 型，只有少数噬菌体(如 T4)具有 Ⅱ 型 tRNA 基因；而真核生物 tRNA 基因均为 Ⅱ 型。然而无论是原核生物，还是真核生物，都有 tRNA 核苷酸转移酶及其基因(*cca*)。*E.coli cca* 基因如果发生突变，则生长十分缓慢。这就表明。tRNA 分子 CCA 末端经常处于核酸酶的作用之下，发生降解，而 *cca* 基因产物的功能不仅在于给 Ⅱ 型 tRNA 加上 CCA3′-OH 末端，而且要负责3′末端发生损伤的 tRNA 分子的修复。该酶的反应是相当专一的。在 ATP、CTP 和缺少 CCA 的 tRNA 存在时，总是生成 CCA 末端，反应需 $Mg^{++}$；ppi 能抑制该反应并使反应逆转。只有当条件改变时，才会出现异常序列的合成。该酶与底物 tRNA 的作用，几乎没有种族特异性。不同种类的酶与 tRNA 的配合都能添加 CCA 末端。然而，用其他类型的 RNA 以及 DNA 作

底物，一般都不能反应。有少数病毒 RNA 能作为底物，可能这些 RNA 包含了该酶所需要的空间结构特征。用改变 tRNA 序列的方法也证明了该酶对于 tRNA 总体结构的依赖性。例如删除反密码子附近的 5 个核苷酸，基本上不影响该酶的作用，而 tRNA5′ 末端除去 5 个核苷酸，则反应能力急剧下降。

RNAaseP：RNAaseP 是一种内切核酸酶。所有的 tRNA5′ 末端都是由该酶产生的，不管 tRNA 序列处于 *trn* 操纵元还是 *rrn* 操纵元的转录产物中。*E. coli* 的 RNAaseP 温度敏感突变体在非许可温度时使 40 多种 tRNA 的前体都发生积累。该酶是一种核蛋白，有两种组分。一种是分子量为 20KDa 的 $C_5$ 蛋白质，另一种为 375 个核苷酸分子量约 120KDa 的 $M_1$RNA。如果用 RNAase 或蛋白酶处理 RANaseP，都可以使酶失活。该酶浮力密度很高，在 CsCl 中达到 1.71g/cm$^3$，这说明 RNAaseP 是以紧密结合的核蛋白形式存在的。重建试验表明，目前没有一种 RNA 分子能代替 $M_1$RNA 与 $C_5$ 蛋白质组成 RNAaseP，也没有一种蛋白质能代替 $C_5$ 蛋白质与 $M_1$RNA 组成 RNAaseP。然而，由于 RNAaseP 两种成分的高度保守性，因此 *E. coli* 的 RNAaseP 成分与人类的 RNAaseP 成分能够交叉重建成为有活力的 RNAaseP。

各种成熟的 tRNA 分子在其 5′ 端都没有序列的同系性。然而用各种 tRNA 的前体分子与 RNAaseP 起反应，都得到该 tRNA 的正确的 5′ 端序列。RNAaseP 为内切核酸酶，产生 5′-磷酸基团和 3′-OH 基团。因此，RNAaseP 识别的是 tRNA 分子的空间构象，特别是 5′ 末端的空间构象，而不是切点附近的几个核苷酸序列。各种 tRNA 都有几个保守的核苷酸，对于维持 tRNA 空间构象是必需的。$M_1$RNA 很可能通过与这些核苷酸产生氢键而识别其切点。该酶的底物专一性很高，只作用于 tRNA 前体；而 rRNA，各种 mRNA 以及成熟的 tRNA 分子都不能作为 RNAaseP 的底物。RNAaseP 与底物作用亦无种族特异性。酶反应需要 $Mg^{++}$ 和一种一价阳离子（$NH_4^+$、$K^+$、$Na^+$），最适 pH 为 8 左右。酶反应能为 tRNA 和 rRNA 所抑制。

RNAaseD：为外切核酸酶，作用于 tRNA3′ 末端多余的核苷酸，直到 CCA 序列为止。一次切除一个，产生 5′ 单核苷酸。根据体外试验，RNAaseD 对于 CCA 的下游序列的活性要高很多，而对 CCA 以及 CCA 上游序列也有一定的攻击能力，尽管其速度低。据推测，在体内由于产生 CCA3′-OH 末端以后立即产生氨酰化，从而保护了 CCA。当 RNAaseP 存在时。RNAaseD 才能发挥其最大活力，可能是由于 RNAaseP 也参与了 3′ 末端的加工。此外，人们还发现了一些作用与 RNAaseD 相当的酶，如 RNAaseQ，RNAaseY，$RNAaseP_3$ 等。不过这些酶都还没有经过很好的鉴定。

RNAaseⅢ 以及相关的酶：这些内切酶负责切开基因之间的间隔子序列。RNAaseⅢ 已在前面讲过。人们还发现了其他类似的酶，如 RNAaseO 和 $RNAaseP_2$。RNAaseO 可能就是 RNAaseD，而 $RNAaseP_2$ 则可能是另一种酶。

$RNAaseP_4$：是一种内切核酸酶，在体外试验中，它能一举切除体外组建的 $tRNA^{Tyr}$ 的前体分子中 CCA 下游的 8 个核苷酸，在体内的情况还不清楚。

## 第八节　转录后处理（三）　内元的去除——RNA 的拼接[72~112]

### 一、概述

RNA 拼接机制的研究，是 80 年代生物化学和分子生物学领域中最有生气的研究课题之

一，它不仅解决不连续基因转录产物的拼接问题，而且对于了解不连续基因的起源乃至整个生命起源与进化问题都是有力的推动。核酸分子的催化功能的发现拓宽了人们对于酶的认识。近几年人们对于RNA拼接机制的研究取得十分惊人的进展。

不连续基因中的介入序列，又称为内元（intron）；被内元隔开的基因序列称为外元（exon）。一个基因的外元和内元都转录在一条原初转录物RNA分子中，然后把内元切除而把外元连接起来，才能产生成熟的RNA分子。这个过程就叫做RNA拼接(RNA splicing)。内元上游方向的一个拼接点称为5′拼接点或左拼接点，内元下游方向的一个拼接称为3′拼接点或右拼接点。Davies等人(1982)建议了"中部核心结构"(central core structure)在内元拼接中的作用，构成这种二级结构的是四个10～12碱基的保守序列。但并非所有的内元都含有这样的二级结构和保守序列。于是Michel等人(1982)将酵母线粒体基因的内元分为两类，大部分酵母线粒体基因的内元都含有上述的二级结构和保守序列，称为第I类内元(group I)，而其余四个内元（啤酒酵母线粒体细胞色素C氧化酶亚基a和亚基b基因的内元aI1，aI2，aI5和bI1)则不具备上述二级结构和保守序列，称为第II类内元(group II)。然而，不久就发现，这两类内元分布很广，特别是第I类内元，并不限于细胞器基因，而且在核基因中也有其代表。特别是纤毛原生动物四膜虫(*Tetrahymena*)的大rRNA前体的拼接机制已经相当清楚，并成为第I类内元拼接机制的模型。第II类内元的拼接机制直到不久前才初露端倪。Peebles等人(1986)和van der Veen等人(1986)均以酵母细胞色素C氧化酶亚基a基因的内元aI5（即I5γ或aI5C)为对象，初步阐明了第II类内元的拼接机制。另一类(是否可以叫第III类?)内元是核基因mRNA前体中的内元，其拼接机制人们早已用β-珠蛋白基因的内元为研究对象，是snRNP(small nuclear RNP)所参与的极其复杂的拼接机制。随着第II类内元的拼接机制的了解以及第II类内元在体外自我拼接的成功，人们必将很快地了解核mRNA前体的拼接机制；并有可能证明RNA自我拼接具有普遍意义。还有一类内元是tRNA基因的内元，均位于tRNA反密码子嗜噗上。含有前三类内元的基因的拼接机制有共同特点，而tRNA前体的拼接则完全不同。四种内元的拼接点附近的序列(边界序列)列于表6-1。

**表6-1 各种内元的边界序列**

（带底线的表示相当保守的序列，箭头表示拼接点，Y代表嘧啶核苷酸）

| 内元类型 | 5′拼接点 | 内元 | 3′拼接点 |
|---|---|---|---|
| 酵母核tRNA内元 | 无保守序列 | 富含A、U | 无保守序列 |
| 第I类内元(线粒体、核) | U↓ | ………… | G↓ |
| 第II类内元(线粒体) | ↓GUG CG | ………… | $Y_n$AU↓ |
| 核mRNA内元 | $^{C}_{A}$AG↓GU$^{A}_{G}$AGU | ………… | $Y_n$AG↓ |
| 酵母核mRNA内元 | ↓GUAUGU | ………… | $Y_n$AG↓ |
| 眼虫叶绿体mRNA内元 | ↓GUG$^{C}_{U}$G | ………… | $Y_n$A$^{U}_{C}$↓ |

## 二、tRNA的拼接

在酵母的约400个核tRNA基因中有40个是不连续基因。每一个都只含有一个内元，恰好位于反密码子的3′方向而且与反密码子相毗邻。内元长度大约为14～46bp。在携带同一

种氨基酸的各种 tRNA 中内元序列是相关的，而不同氨基酸的 tRNA 的内元序列是不相关的。tRNA 基因的内元没有为拼接酶系所识别的一致序列。所有的 tRNA 内元都有与其反密码子互补的序列。因此，在拼接之前的前体中反密码子总是处于双链 RNA 状态，从而受到保护而不致为拼接酶系所破坏。这部分双链使得反密码子噜噗的柄部比成熟 tRNA 的反密码子噜噗的柄部长得多(图 6-27)；而三叶草结构的其余部分与成熟的 tRNA 完全一样。因此，拼接酶系识别的是共同的二级结构，而不是共同的内元序列。

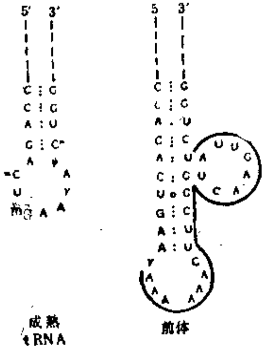


图 6-27 酵母的成熟 $tRNA^{Phe}$ 前体分子及其前体反密码子噜噗(画线部分为内元序列，粗体 GAA和 GAA 为反密码子)

从酵母 $tRNA^{Tyr}$ 及其两种突变 tRNA 前体拼接情况(图 6-28) 来看，内元部分的不同配对状况产生了不同稳定性的双链 RNA。这种状况影响拼接效率，而不影响拼接的精确性。由此看来，内元部位只是拼接酶的作用位点，而其识别位点可能是整个 tRNA 分子的构象或者 tRNA 分子中某些保守的结构特征。与此相一致的是，成熟的 tRNA 能竞争性地抑制酶对于 tRNA 前体的拼接。

有一种温度敏感突变体在较高温度时不能拼接不连续 tRNA 基因，而非不连续基因所编码的 tRNA 可以照样成熟并转运到细胞质中。这样，人们就可以研究拼接反应的机制。将较高温度下生长的突变体中积累的 tRNA 前体分

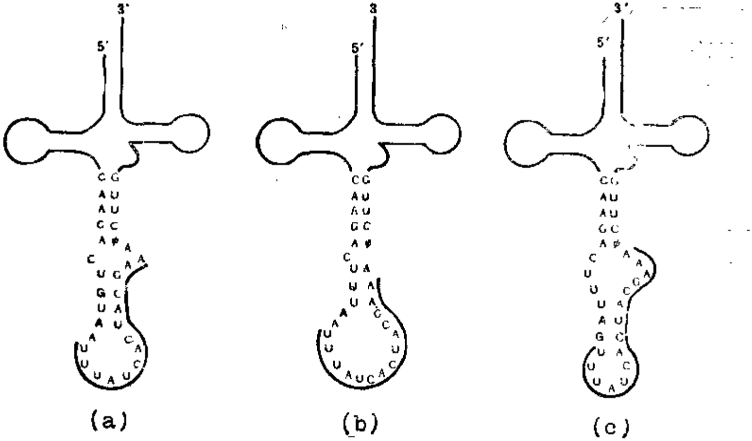
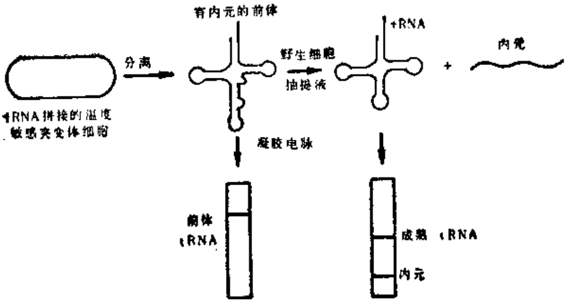


图 6-28 酵母 $tRNA^{Tyr}$ 及其突变体的前体分子的拼接状况

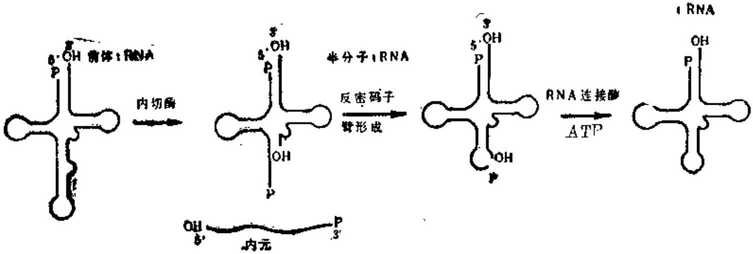
(画线部分为内元序列，粗体字母表示反密码子)

(a) $tRNA^{Tyr}$ 前体分子中，两个拼接点都处于不配对的单链部分，能正常拼接。(b)Sup4 前体分子中，左拼接点处于单链部分，右拼接点处于双链部分，拼接亦能进行。 (c) a 122 突变体中，左拼接点处于双链部分，右拼接点处于单链部分，拼接效率很低。

离出来，然后加入野生型细胞抽提液，即可发现前体发生拼接的现象(图 6-29A)。如果控制无细胞抽提液中的 ATP，则可以发现整个反应为两步：第一步是由一个内切核酸酶所催化的两个磷酸二酯键的断裂，释放出一条线状内元分子和两个相互由二级结构作用力连接在一起的所谓"tRNA 半分子"(tRNA half-molecules)。这一步反应不需要 ATP，催化这步反应的酵母内切核酸酶表现出膜结合的性质。tRNA 半分子很快采取成熟 tRNA 分子的构象，只不过在反密码子环上留下一个切刻。然后再发生第二步反应，即由 RNA 连接酶所催化的依赖于 ATP 的连接反应(图 6-29B)。




(a)




(b)

图 6-29

(a) 突变体中积累的 tRNA 前体分子可被野生型细胞抽提液拼接；

(b) 拼接过程中第一步反应不需要 ATP，第二步反应依赖于 ATP。

从图 6-29B 可以看出，去除内元的内切核酸酶与其它参与 tRNA 转录后处理的酶不同，在每个切点上断裂磷酸二酯键后产生 5'OH 和 3' 磷酸，而 3' 磷酸很快转变为2'，3'-环式核苷酸。这样，生成的 tRNA 半分子，一个含有两个磷酸末端，一个含有两个-OH 末端。在酵母中，连接过程包含了一系列复杂的生化反应。第一个外元的 3' 末端的环式核苷酸在环式磷酸二酯酶作用之下开环，使 3' 成为 OH。而第二个外元的 5' 端在多聚核苷酸激酶的作用下，OH 转变为磷酸基。这样，RNA 连接酶才能将两个外元连接起来。最后切除第一个外元的 2'-磷酸(这是 tRNA 内元拼接的重要特点之一)，从而完成了 tRNA 的拼接(图 6-30)。拼接酶系对于底物要求没有种族特异性。将酵母 tRNA 前体用非洲爪蟾卵母细胞核的抽提液进行拼接

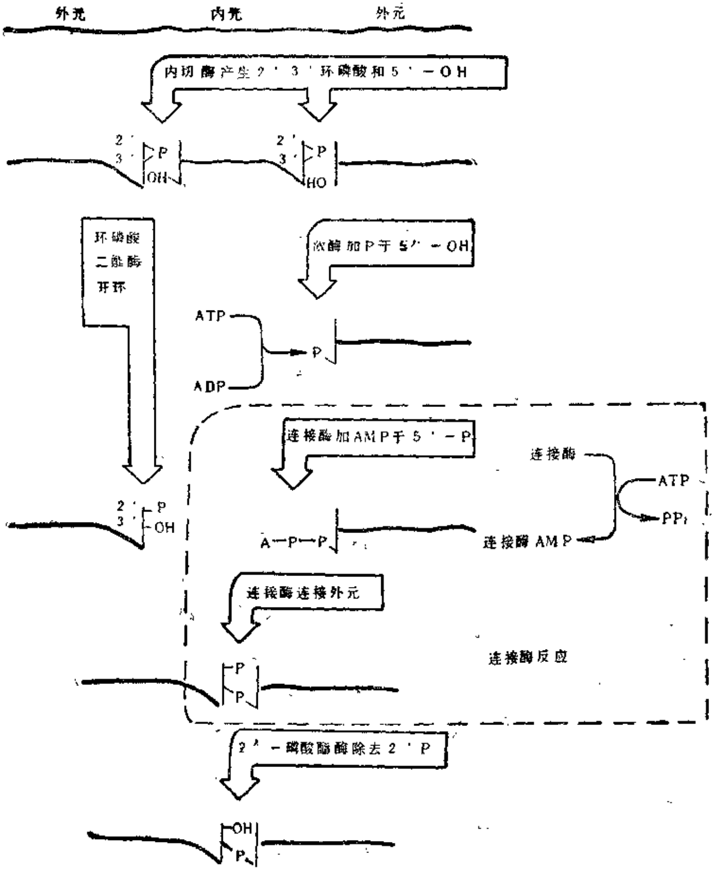



图6-30　酵母tRNA半分子的连接过程

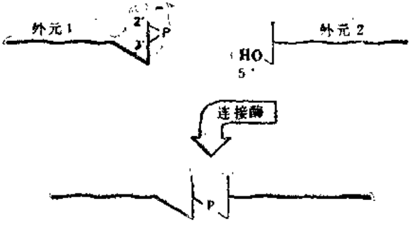



图6-31　哺乳动物RNA连接酶可能直接将2',3'-环式磷酸基团与5'-OH连接起来

也同样奏效。人类 Hela 细胞中的 RNA 连接酶能直接连接带有 2′,3′-环式磷酸基团的第一个外元和带有 5′-OH的第二个外元。形成新的磷酸二酯键的磷酸基团仍然是原来存在于第一个外元末端的那个磷酸基(图 6-31)。

### 三、第Ⅰ类内元的拼接

第一类内元,其 5′ 拼接点和 3′ 拼接点的序列绝大部分为 U↓……G↓,例如酵母(啤酒酵母)细胞色素 b 基因(Sc COB),细胞色素 c 氧化酶亚基 a 基因(Sc OX),rRNA基因(Sc rRNA)以及四膜虫 rRNA 基因(Tt rRNA)等基因中的内元的两个拼接点序列如图 6-32 所示。

| | | | |
|---|---|---|---|
| Sc COB3 | 5′GUU↓UAAUA | …… | UAAAUG↓CUC3′ |
| Sc COB4 | 5′GGU↓CAAAA | …… | UAAAAG↓CAU3′ |
| Sc COB5 | 5′UAU↓UGAUA | …… | AAAUUG↓UGU3′ |
| Sc OX3 | 5′GGU↓AACCA | …… | AAAAUG↓AAC3′ |
| Sc OX4 | 5′GGU↓CAAAC | …… | AACAAG↓CAC3′ |
| Sc OX5α | 5′AUU↓AAUAA | …… | UAAAAG↓AGC3′ |
| Sc OX5β | 5′GAU↓AUUAA | …… | UUUAUG↓ACU 3′ |
| Sc rRNA | 5′GAU↓AAUUU | …… | AAUUUG↓AAC3′ |
| Tt rRNA | 5′UCU↓AAAUA | …… | UACUCG↓UAA3′ |

图 6-32 部分第Ⅰ类内元的拼接点序列

除了拼接点序列特征之外,第Ⅰ类内元还具有比较保守的四种 10～12 核苷酸的序列,分别以 5′-P-Q-R-S-3′ 表示。P 序列能与 Q 序列互补,R 序列能与 S 序列互补,形成一个所谓中部核心结构。这样,位于内元序列中靠近 5′ 拼接点的一段序列就能与两个拼接点的边界序列配对,将两个拼接点拉在一起便于磷酸二酯键的切断和再连接。内元中能与两个拼接点边界序列配对的一段序列称为内部引导序列(internal guide sequence,简称 IGS)。P、Q、R、S 的一致序列示于图 6-33。四膜虫 35S RNA 中的内元的二级结构示于图 6-34,其中包括中部核心序列 P,Q,R,S,9R′,9R 以及内部引导序列与 5′ 边界序列的配对情况。

| | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | 11 | 17 | 13 | 16 | 10 | 12 | 14 | 13 | 18 | 11 | | |
| P: | A | U | G | C | U | G | G | A | A | A | | |
| | U | | | G | A | A | A | G | | U | | |
| | 7 | | | 3 | 6 | 5 | 5 | 5 | | 5 | | |
| | 14 | 19 | 15 | 11 | 9 | 16 | 12 | 17 | 16 | 10 | | |
| Q: | A | A | U | C | A | G | C | A | G | G | | |
| | U | | C | U | U | C | | | | C | | |
| | 5 | | 3 | 5 | 7 | 3 | | | | 7 | | |
| | 16 | 16 | 18 | 10 | 10 | 19 | 19 | 18 | 19 | 16 | 17 | |
| R: | U | C | A | G | A | G | A | C | U | AX | A | |
| | | U | | A | C | | | | | | | |
| | | 3 | | 6 | 6 | | | | | | | |
| | 19 | 16 | 17 | 16 | 13 | 18 | 16 | 19 | 18 | 19 | 18 | 11 |
| S: | A | A | G | A | U | A | U | A | G | U | C | C |
| | | U | | G | | | | | | | | U |
| | | 3 | | 3 | | | | | | | | 6 |

图 6-33 构成核心结构的四个保守序列,底线表示 P-Q组互补或 R-S 组互补的核苷酸。共分析 19 个内元序列。

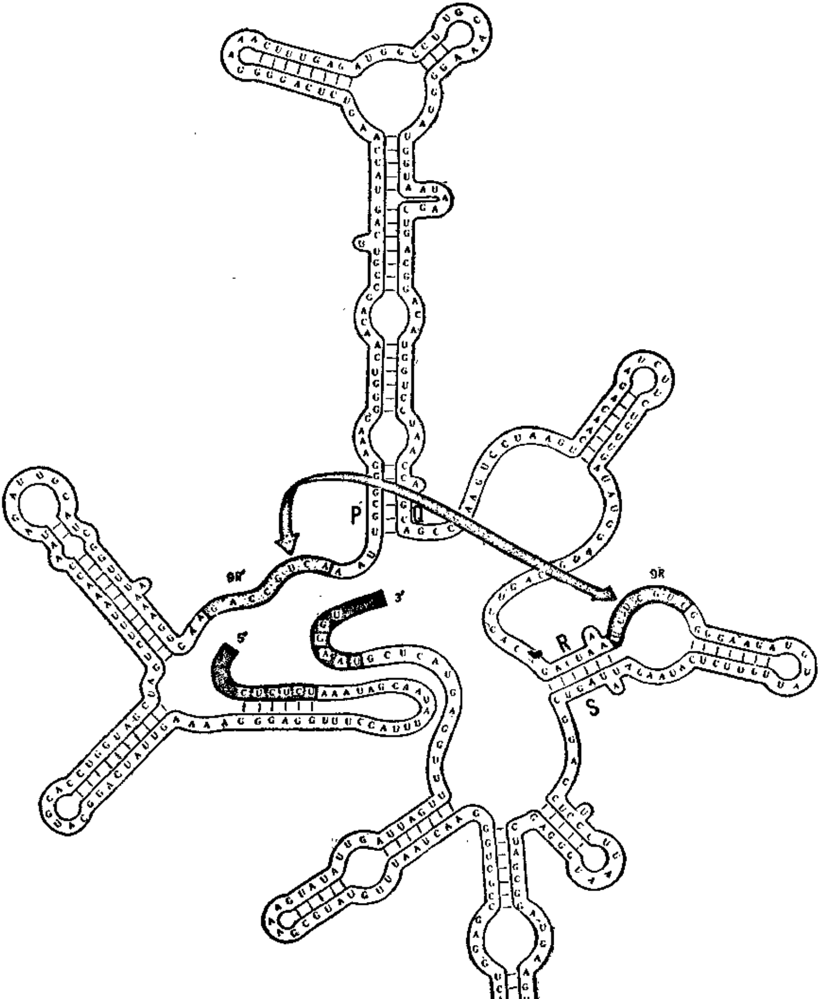



图 6-34　四膜虫 35S RNA 中内元的二级结构，在其三维结构中，9R′ 与 9R 两个序列也是配对的。其内部引导序列为 3′GGGAGGUUUCCAUU5′，外元序列只写出靠近边界的各 6 个核苷酸

下面，我们以四膜虫的大 rRNA 前体分子的拼接为例，说明第 I 类内元拼接的分子机制。

在四膜虫的生活史中有一个阶段，在这个阶段中二倍体基因组保留在微核中，而基因扩增发生在大核中。rDNA 的扩增产生了大量的染色体外线状分子，每个分子都是迴文序列的二聚体。这个二聚体含两个完全相同的转录单元，转录成为 35S 的 rRNA 前体。含有内元的 26S rRNA 位于这个前体的 3′ 端部分。

将分离到的大核进行保温，35S RNA 则释放出 413 核苷酸长的线状 RNA 单链，即内元序列，而没有游离的外元出现。如果延长保温时间，则释放出来的线状内元变成一个环形单链分子(图 6-35)。将分离得到的 35S 前体分子在一种只含一价阳离子，二价阳离子和 GTP溶液中时，能得到同样的结果。用 GDP、GMP 和鸟嘌呤核苷也同样可以，而只要该核苷(酸)含有 3′-OH。这就表明，这种反应并不需要能量输入。其他种类的核苷(酸)都不能代替鸟嘌呤核苷(酸)。

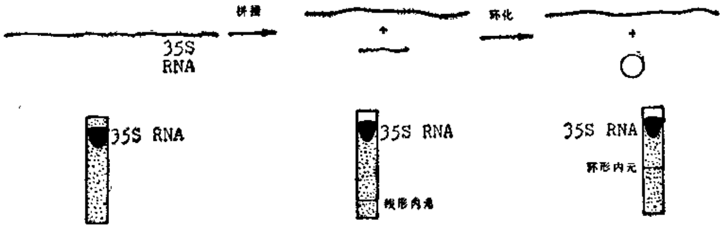



图 6-35　通过电泳检定四膜虫 35S rRNA 前体的拼接过程

用放射性标记的鸟嘌呤核苷处理 35S RNA，则放射性进入线状的内元片断中。现已查明，G通过正常的磷酸二酯键与内元序列的5′端相连。这是鸟核苷的3′-OH进攻第一个外元与内元之间的磷酸二酯键的结果，同时在第一个外元 3′ 末端生成了羟基。然后，第一个外元的 3′-OH 进攻内元与第二个外元之间的磷酸二酯键，结果两个外元序列连接起来。而释放出来的线形内元序列再经过两次连续的磷酸酯转移反应，释放出 15 核苷酸和4核苷酸两个小片断，最终成为稳定的线形产物，称为 L-19(图 6-36)。在转酯反应中，由一个磷酸酯键转变成另一个磷酸酯键，因此破坏的磷酸酯键与生成的磷酸酯键总是相等的。值得注意的是在转酯反应过程中，不需要任何额外能量的输入。第一次转酯反应和第二次转酯反应必然是紧密耦联的，因为观察不到自由的外元序列的存在。同样，内元的释放和外元的连接也必然是同一反应的两个结果。

上述的 L-19，是四膜虫 35S RNA 拼接的最终产物。在内元从原初转录产物上切除下来后，内元又进行了两次转酯反应，其意义可能在于使基因拼接能不可逆地进行下去。这一拼接反应在没有任何蛋白质存在的情况下就能够进行，也就是说这种活性是 35S RNA 的内元本身所具有的性质，因此叫做自身催化(auto-catalysis)。这种自身催化是将一处已有的磷酸二酯键转变成另一处新的磷酸二酯键，而不需要以高能的前体(NTP)的分解为代价合成新的磷酸二酯键。从这一意义上来说，内元可以看作是RNA 重排酶(RNA rearrangease)或 RNA 异构酶(RNA isomerase)。

然而，在拼接反应中，“酶”和底物是同一个分子，而且这个“酶”在反应结束后并没有再生，而是变成了不同的分子。因此人们曾经对于四膜虫 35S RNA 的内元是不是酶的问题产生了疑问。美国科罗拉多大学 Cech 等人进行了出色的工作。他们的工作一方面阐明了四膜虫

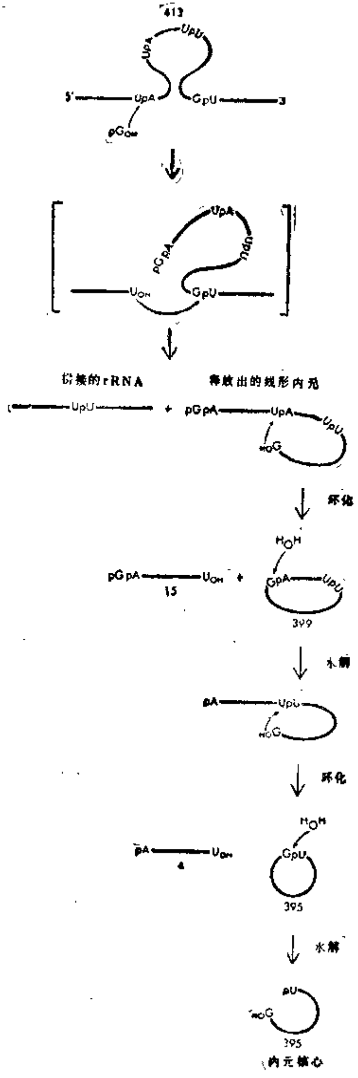



图 6-36 四膜虫 35S RNA 拼接的化学本质是连续的磷酸酯转移反应

35S RNA 的拼接机制，另一方面又证明了 L-19 分子具有 poly C 聚合酶活性。内元序列折叠为一定的空间结构，其内部引导序列又与边界处的外元序列配对。当鸟核苷与其结合位点(内元序列中某种特定的空间构象)通过氢键相结合时，其 3′OH正好处于向 5′ 端外元的最后一个核苷酸的 3′ 磷酸酯键发动亲核进攻的位置上，于是 5′ 拼接点发生断裂。与此相类似，当 5′ 拼接点断裂之后，5′ 外元的 3′ 羟基也正好处于向 3′ 拼接点进攻的有利位置，同样也需要内部引导序列的作用。这一作用机制示于图 6-37。当 L-19 分子与寡聚 C(如 $_pC_pC_pC_pC_pC$)保温时，L-19 能够催化 polyC 的生成。Cech 用这种方法已得到有数十个核苷酸的 polyC。这一过程示于图 6-38。这里，人们不难得出结论：L-19 具有酶的主要特征，专一性强，加快反应速度，反应前后酶分子保持不变。这种由 RNA 构成的酶，人们称之为核糖核酸质酶(ribozyme)，以别于蛋白质酶(proteineous enzyme)。为了方便起见，本书有时将核糖核酸质酶简称作 R 酶。

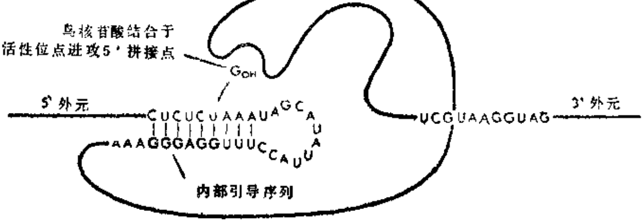



图 6-37　四膜虫 35S RNA 的内元采取一定的空间构象，内部引导序列与边界序列配对使得转酯反应能够顺利而正确地进行。图中只显示鸟核苷的 3′-OH 对于 5′ 拼接点的进攻态势

尽管这种催化性质是 RNA 本身的性质，但在体内可能有某些蛋白质能稳定RNA的结构。从而协助了 RNA 的催化作用。RNAaseP 就是一个突出的例子。在体内只有当 RNA 组分 $M_1$RNA 和蛋白质组分 $C_5$ 蛋白质同时存在时才能起作用，然而在体外，只要增加 $Mg^{++}$ 浓度，就使得 $C_5$ 成为多余的了。$M_1$ 本身就能够起催化作用，而且与 RNAaseP 的作用差不多。由此可见，这种 RNA 具有酶的性质。在 *E.coli* 中，$C_5$ 基因的突变和 $M_1$ RNA 基因的突变都可使 RNAaseP 失活。人们根据习惯都把蛋白组分看成是负责催化作用的成分，而 RNA 则看成是辅助成分，例如能通过碱基配对促进与底物的结合。现在已经弄清，这样作用应该颠倒过来。

RNA 为什么能成为催化中心？催化中心无非是提供一个表面，让参加反应的基因靠近并形成固定不变的空间关系，蛋白质以其侧链基因的多样性构成现代生物细胞中绝大部分的酶。然而，RNA 分子也能够以其分子的空间构型产生键的断裂和生成所必须的环境。催化 RNA 和底物 RNA 之间的碱基配对可能是产生这种催化环境的主要原因(图 6-39)。

RNA 催化功能的发现对于了解生命进化过程有着重大的意义。很可能在原始的生命中，RNA 所催化的断裂-连接反应是最早出现的催化过程。根据从遗传信息到蛋白质合成的过程中 RNA 的多方面作用(mRNA、tRNA、rRNA 等)，人们早就推测 RNA 是生物体系中最早出现的大分子种类(Crick，*J. Mol. Biol.*，**38**：367～379，1968)。酵母 tRNA 拼接过程中所涉及的 2′ 磷酸酯键，是 DNA所不可能有的。因此 RNA 催化功能必然出现在 DNA 成为遗传信息载体之前。也就是说，在原始生命中遗传信息载体是 RNA 而不是 DNA。由于 DNA 的稳定性和空间结构方面的某种优势而取代了 RNA 成为主要的遗传物质；而蛋白质以其侧链的多样

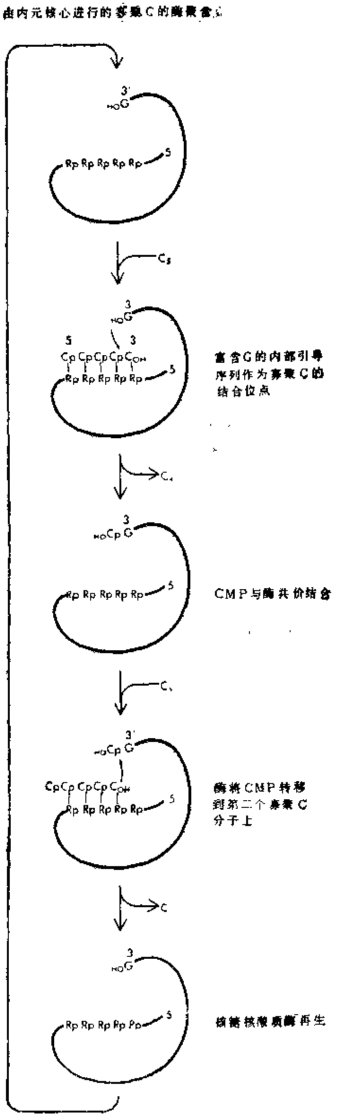



图6-38　四膜虫35S RNA内元核心L-19分子的poly C聚合酶活性。注意图中3′末端的G-OH不是外加的鸟苷，而是内元3′末端的鸟苷酸

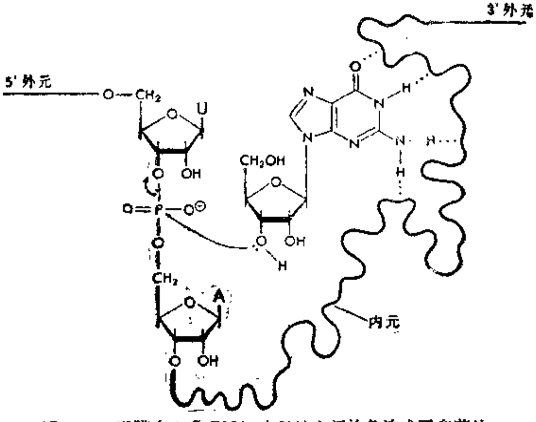



图 6-39　四膜虫 35S RNA 内元的空间构象造成了鸟苷的3′-OH向5′拼接点亲核进攻的环境。这与牛胰核糖核酸酶激活2′-OH去进攻相邻的3′位上的磷酸酯键的情况十分相似。

性和灵活性而取代 RNA 成为主要的催化剂(酶)。

RNA 的催化功能可能负责原始体系中的 RNA 复制。P.A. Sharp 提出了这种不依赖于蛋白质的复制体系假说。双链 RNA 在变性时，其中一条链成为催化 RNA，另一条链成为模板。热能或某种原始聚合活性促进了在模板链上形成较短的核苷酸片断。催化 RNA 能够修

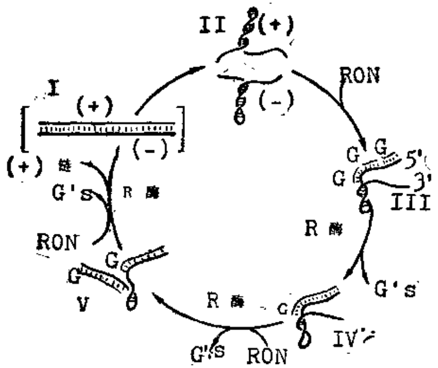



图 6-40　RNA 自我复制的模式图。双链 RNA (I)经受变性和复性的变化，其正链在变性后能折叠成具有催化功能的 R 酶(Ⅱ)。带有5′-G 的随机寡聚核苷酸(RON)装配在模板上，按 Watson-Crick 原理配对(Ⅲ)。R 酶修剪去多余的核苷酸并将各寡聚核苷酸连接起来(IV)。这一过程继续进行(V)。最后得到双链 RNA(I)。这一双链 RNA 是否作为中间产物而存在，现在还不能肯定。

别这些寡聚核苷酸片断的相互重叠部分以及多余的非同系部分，然后加以连接。Cech和他的同事对于四膜虫 rRNA 的内元的催化性质作了大量的研究。1989 年，J.A.Doudna 和 J.W.Szostak 在独立的模板上，将寡聚核苷酸与模板的碱基配对，然后用四膜虫 rRNA 内元这个核糖核酸质酶将这些寡聚核苷酸修剪并连接起来(注：这里所使用的寡聚核苷酸除与模板全部配对外，在每个接口的 5′ 端都多一个鸟嘌呤核苷酸)。根据这些实验和 Sharp 的假说，Cech 提出了 RNA 自我复制的模式图(图 6-40)。根据这一模式，在只含 RNA 和寡聚核苷酸(或单核苷酸)的体系中就能完成 RNA 复制。催化 RNA 也同样作为模板指导其互补链的复制。

RNA 自我复制和自身拼接是极其古老的机制，具有这种催化功能的核糖核酸质酶——不连续基因中的介入序列(内元)也就必然是一种古老的核酸序列。当 70 年代末刚发现内元时，人们就认为它是真核基因的真质标记(hallmark)。1984 年，Chu 等人发现了 T4 噬菌体的胸腺嘧啶核苷酸合成酶的基因中有一个 1 017 个核苷酸的内元。这就推翻了真质标记说。

T4 内元符合第Ⅰ类内元的特征，但又包含其特殊性。其内元以 Ochre 终止密码子 UAA 开始，以起始密码子 AUG 结束(图 6-41)。这就提出一个问题，这个内元序列到底是在 RNA 水平上拼接除去的呢还是在翻译水平上翻译成两个肽段然后再连接起来的呢？杂交试验表明，在 T4 感染的 *E.coli* 中发现三种 T4 编码的 RNA：2.5Kb 的 RNA 前体、1.3Kb 的成熟 mRNA 和 1Kb 的内元 RNA。而且，人们并未分离到与外元相应的两个肽段。因此 T4 内元的删除机制仍然是 RNA 的自我拼接。

外元Ⅰ┈┈┈GGU ↓ |UAA┈┈内元┈┈AUG| ↓ CUA┈┈┈┈外元Ⅱ

图 6-41　T4 胸腺嘧啶核苷酸合成酶基因的原始转录物结构

在 RNA 自我拼接的反应中，除了 RNA 的自身催化作用之外，有时还需要蛋白质因子维持 RNA 的空间构象。这些蛋白质中，有的竟是内元本身所编码的。

### 四、编码 RNA 成熟酶的内元

在许多第Ⅰ类内元中，存在着外元和内元联合编码一种调节蛋白质——RNA成熟酶(RNA maturase)的情况。最近，也发现第 Ⅱ 类内元与外元联合编码成熟酶的情况(如眼虫的叶绿体有 32KDa 的 RNA 成熟酶)。下面我们以酵母线粒体细胞色素 b 基因(*cob* 基因，又称 BOX基因)为例，说明 RNA 成熟酶的生成及其调节作用。

不同的酵母菌株含有长短不同的 *cob* 基因。有的菌株的 *cob* 基因是所谓“长基因”，编码序列 1 155bp，整个基因长大约 6 400bp，包括 6 个外元(B1～B6)和 5 个内元(I1～I5)。有的菌株为“短基因”，大约只有长基因的一半，包括 3 个外元和两个内元。第一个外元相当于长基因的外元 B1～B4，而长基因的内元 I1、I2和I3 在短基因中全部消失。长基因和短基因都编码同样的 385 氨基酸，而且具有同样的表达能力。长、短 *cob* 基因的结构示于图 6-42。

分离 *cob* 基因的 mRNA，可以得到几种不同大小的分子。最短的大约 3 000 个核苷酸，即成熟的 mRNA，包括了编码区和 5′、3′ 端的非翻译区。最长的 8 500 个核苷酸，是未经处理的原始转录物。其他介于二者之间，但大小是不连续的，说明其中有的内元除去了，有的内元仍然留在里面。

妨碍细胞色素 b 合成的突变成簇分布，每簇大约占据一个图距单位，各簇之间相隔2～9个图距单位。这些簇可以分为三类：

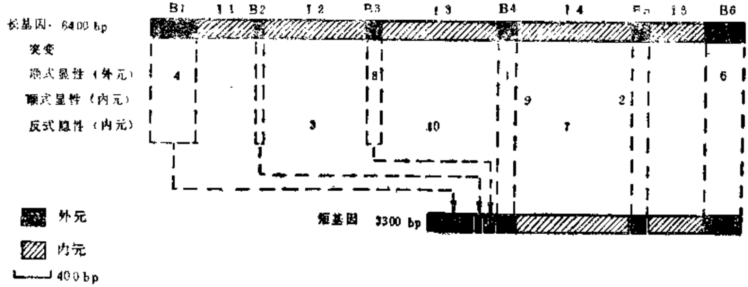


图 6-42 酵母线粒体细胞色素 b 的长基因和短基因的结构。在长基因中突变位点簇的数字是突变体的分离时间顺序，不是基因上的位置次序。

(1) 顺式显性：包括Box4、8、1、6。这些突变体均能转录出mRNA，但是产生了错义突变或无义突变。这四个框里各种突变体相互之间都不能互补，这就说明它们都位于同一个基因中。这四个框分别是外元 B1、B3、B4、B6。由于 B2、B5 很小(B2 = 14bp、B5 = 51bp) 尚未发现其中的突变体。这种突变点的成簇分布是与不连续基因的序列相应的。

(2) 另一种顺式显性：即 Box9 和 2 的突变，其中突变体之间也不能互补，而且也不能与第一类突变体互补。然而它们与第一类突变体不同，它们阻碍 mRNA 的生成。序列分析表明，它们都位于 I4 内，Box9 在离开 I4 左拼接点 350bp 的 8bp 长的序列中，Box2位于 I4 右拼接点启上游方向 25bp 处。两组突变均妨碍 I4 的去除。这组突变，说明了两个问题：内元中特定位点的突变，尽管离开拼接点相当的距离，仍然会影响拼接；这类突变与蛋白质本身的突变无法用经典遗传学方法(即互补试验)加以区别(典型的顺式作用类型突变体，如启动子突变和操作子突变也是如此)。

(3) 反式隐性：即 Box3、10、7 三组突变，它们构成三个互补群，虽然每个群内突变体之间不能互补，但能与不同群的突变体以及上述第一类外元突变体产生互补。这就意味着，这三组突变中任何一组，都编码一个反式作用的可扩散的蛋白质，它可能通过某种调节作用影响到细

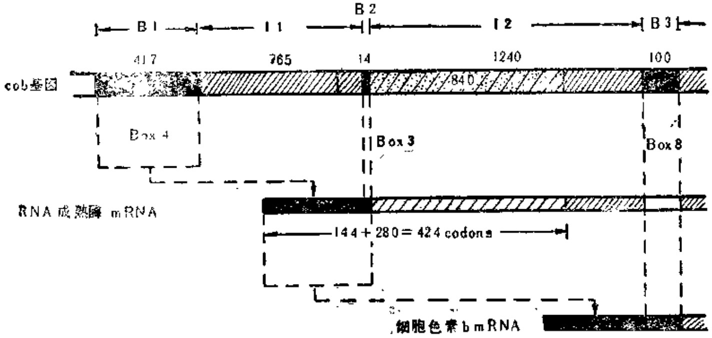



图 6-43 *cob* 基因 RNA 的连续拼接能够产生序列重叠的蛋白质(RNA 成熟酶)的mRNA

胞色素 b 的合成。这三组突变均阻止成熟 *cob* mRNA 的产生，结果导致了不同大小的mRNA前体的积累：Box3 突变积累7 500b，Box10 突变积累 7 100b，Box7 突变积累 3 500b的mRNA前体。说明这三组突变体所影响的产物是在拼接的不同阶段发生作用的。序列分析结果表明，Box3 位于 I2，Box10 位于 I3，Box7 位于 I4(图 6-42)。原来，*cob* 基因中 I3，I4，I2 各编码一个为删除各自的内元序列所不可缺少的蛋白质。这类蛋白质就叫做 RNA 成熟酶，其作用在于稳定 RNA 的空间构象，而不是催化作用。

图 6-43 以内元 2 所编码的 RNA 成熟酶为例，说明成熟酶 mRNA 的产生过程。第一个外元(B1)编码细胞色素 b 的 N 端 139 氨基酸(417bp)。第一个内元(I1)765bp，其中三个阅读框均被阻断。B2 是一个很小的外元，只有 14bp。I2 长 1240 bp，在其前 840bp，有一个开放的阅读框，与 B1+B2 的阅读框完全一致。这样，当 I1 被除去时，B1，B2 与 I2 及其以后部分就连接在一起，成为一个较短的 mRNA 前体。这样，从 B1 的 AUG 密码子开始，共有 144+280=424 个密码子，能翻译产生一种 423 个氨基酸的 RNA 成熟酶。其前 144 个氨基酸与细胞色素

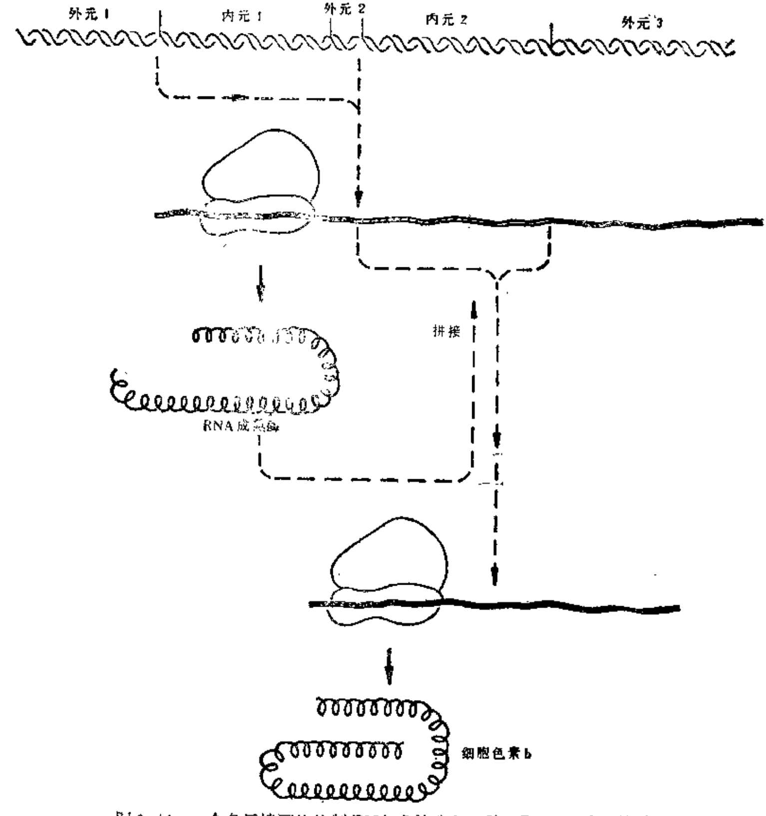



图 6-44 一个负反馈网络控制 RNA 成熟酶和两种 mRNA 形式的数量

b 的 N 端 144 个氨基酸完全相同，而其余 279 个氨基酸则由 I2 编码。

这个 RNA 成熟酶是去除 I2 本身所不可缺少的。毫无疑义，当 I2 去除之后 B1、B2、B3 相互衔接起来，就不再产生这个 RNA 成熟酶。这样，就构成了一个自动调节的负反馈网络(图 6-44)。正是由于 RNA 成熟酶导致自身合成能力的丧失，因此细胞内 RNA 成熟酶含量极低。线粒体内存在这一种负反馈机制，就说明一种 RNA 成熟酶只作用于其相应的内元的拼接过程，而对于其他内元的拼接不起什么作用，至少大多数情况是如此。但是人们发现，*cob* 基因的 I4 突变体(Box7 突变)不但妨碍 *cob* mRNA 的拼接，而且妨碍 *oxi3* 基因 mRNA 的拼接。Box7 突变在 I4 的开放阅读框中产生一个无义密码子，从而中断了 I4 (及其以前的外元序列)所编码的 RNA 成熟酶。这个酶是去除 I4 和 *oxi3* 的某个内元所必需的。由于细胞色素 b 的作用和细胞色素氧化酶的作用是紧密相关的，这也可以看作是一种协调机制。

酵母线粒体基因的内元不具备核 mRNA 基因内元那种 GT-AG 序列，也不具备一致序列，因此编码 RNA 成熟酶对于其自我拼接是很重要的。RNA 成熟酶可能负责内元的识别，构成拼接复合体中的一个特异性亚基来维持 RNA 的空间构象并促进 RNA 的自我拼接。

### 五、第 Ⅱ 类内元的拼接

第 Ⅱ 类内元除了表 6-1 中所示的两个拼接点序列特征之外在其 3′ 末端还有值得注意的序列特征：

(1) 离开 3′ 拼接点 6～12bp 的序列比较保守，一致序列为 $P_yP_uP_yP_yT\overset{*}{A}P_y$，其中 $\overset{*}{A}$ 为百分之百保守的一个核苷酸；

(2) 在上述 7 核苷酸序列的中间和两侧各有一段 3～5 核苷酸的短序列能与上游方向的核苷酸互补；而那个保守 $\overset{*}{A}$ 总是不包含在这些互补的序列之中。也就是说，在形成部分双链

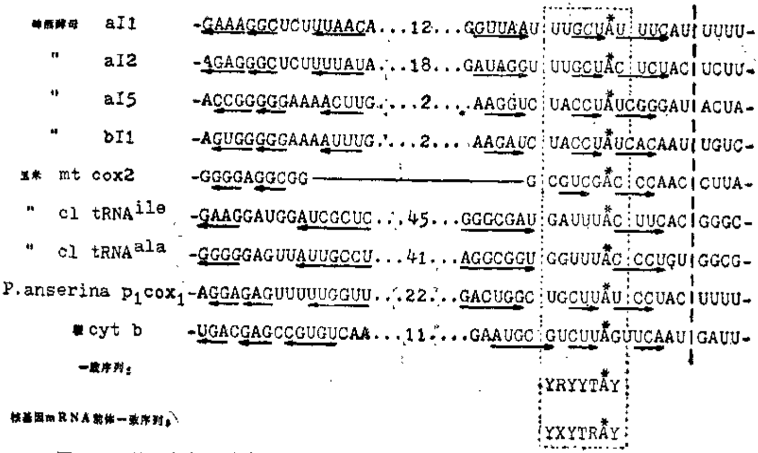



图 6-45 第Ⅱ类内元 3′ 末端的序列特征。图中虚线为3′拼接点；点线框内为保守的 7 核苷酸序列，$\overset{*}{A}$ 为其中全保守的一个核苷酸；中间的数字表示未写出来的核苷酸数；在这两侧带有箭头的序列表示能互补形成 RNA 的二级结构。一致序列中，R 代表嘌呤，Y 代表嘧啶，X 为嘌呤或嘧啶。

结构中，保守 A 总是不在双链部分，而是处于向外突出的"芽状物"之中。上述这些结构特征见图 6-45。

由于 mRNA 前体极不稳定，用传统的从细胞中分离提取 mRNA 前体方法去研究mRNA前体的拼接机制，常常不能得到令人满意的结果。自从 1984 年哈佛大学的 Green 实验室利用高效的噬菌体 SP6 的启动子组成人工质粒以后，RNA 拼接的研究取得了长足进展。最近，Peebles 等人(1986)，van der Veen 等人(1986)将细胞色素 o 氧化酶亚基的内元 I5γ 克隆到包含 SP6 启动子的质粒(如pKM2)中(图 6-46)，则可得到大量的包含此内元的 RNA 转录产物。这种产物具有体外自我拼接的能力。这种拼接与第 I 类内元不同，不需要带有 3′-OH

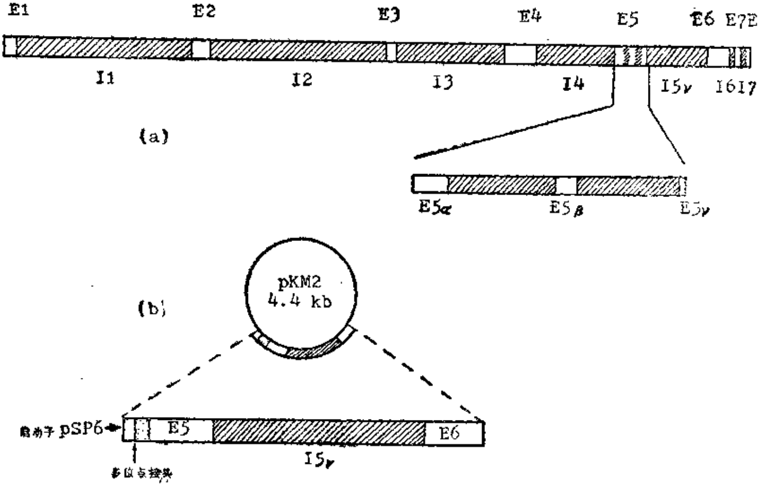



图 6-46 (A) coxI3 基因结构；(B) 含有 I5γ 的 DNA 序列与 SP6 启动子克隆到质粒中去

的鸟核苷或鸟核苷酸，也不需要一价阳离子，仅需要 5mM 左右的 $Mg^{++}$ 和少量亚精胺；如果增加 $Mg^{++}$ 浓度，亦可不用亚精胺。该反应有个很狭的最适温度范围约 45℃ 左右。切下来的内元呈套索状(lariat)，由 2′,5′-磷酸二酯键连接成环。套索状产物如用 HeLa 细胞提取液(含有去分枝酶)处理则可变为线形分子。以套索状产物为模板进行反转录，则在分枝处迅速终止反应。由于套索状结构具有自由的 3′-OH 末端，因此可以用与套索互补的单链 DNA 为模板，以套索为引物合成 DNA 链。两种反应结束后都可以用 RNAaseI 切除套索结构。经过变性、电泳，即可发现新合成的一段 DNA(图 6-47)。在体外拼接过程中，还发现少量只切未连的 5′ 外元序列和 3′ 外元序列，这可能是由于体外体系中缺少维持 RNA 构象的蛋白质因子。

第 II 类内元的自我拼接与第 I 类内元的自我拼接一样，都不需要额外的能量输入，这意味着同样可能是一连串的转酯反应。但第 II 类内元的自我拼接不需要鸟核苷或鸟核苷酸来发动第一个转酯反应。现已证明，发动这第一个转酯反应的正是 7 核苷酸序列中的保守 A。A 以其 2′-OH 发动亲核进攻，保守 A 不参加二级结构的碱基配对，保证了它发动转酯反应的功能。这一连串的转酯过程示于图 6-48。人们推测。核基因 mRNA 的内元的拼接过程是与第 II 类内元的拼接过程相类似的。

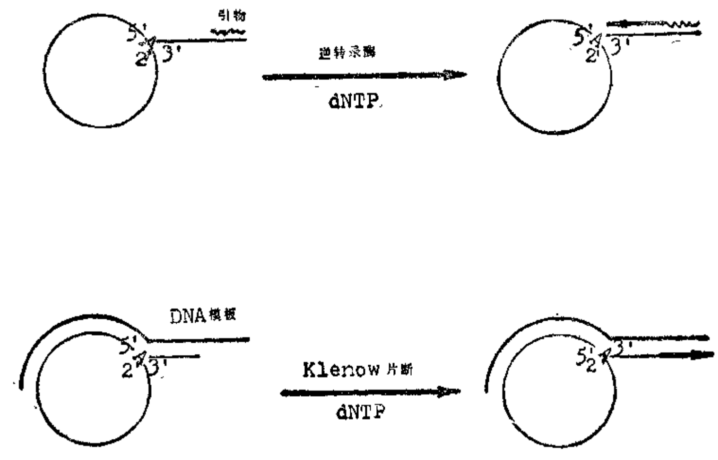



图 6-47　鉴别套索结构的两种方法

(A) 反转录法　(B) klsnow 片断合成 DNA 法

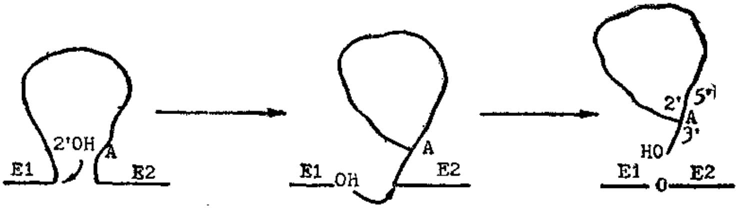



图 6-43　由 7 核苷酸序列的保守 A 的 2′-OH 发动亲核进攻，
产生一系列的转酯反应

## 六、核基因 mRNA 内元的拼接

核基因 mRNA 的内元的拼接点序列很简单，均为↓GU………AG↓，这就是普遍适用的所谓 Breathnach-Chambon 规则。在内元的 3′ 末端，也有一个核苷酸序列，其一致序列是 $P_yXP_yTP_u\overset{*}{A}P_y$。除此以外，再也没有发现其他保守的序列，即使在某一物种或某一基因族内也都没有发现。这两个保守的拼接点序列确实很重要，如果这两个保守序列突变就不能正确地拼接 mRNA 前体。地中海贫血病就是这方面的例子。然而，单靠这样简单的保守序列是不可能做到正确拼接的。现在的普遍看法是 snRNA 参与了这一识别作用。

真核生物的细胞中，有许多 100～300 碱基的小分子 RNA，每个细胞含有 $10^5$～$10^6$ 个分子。它们有的是 RNA 聚合酶 Ⅲ 转录的，有的是 RNA 聚合酶 Ⅱ 转录的，其中有的还像 mRNA 一样戴上帽子。在核中的小 RNA 叫做 snRNA，在细胞质中的叫 scRNA。一般情况下它们都是以核蛋白形式存在的，即 snRNP(snurps) 和 scRNP(scyrps)。在各种 snRNA 之间有

序列的共同性，各种 snRNP 也可能有共同的蛋白质亚基，因为一种 snRNP 的抗体能识别许多种 snRNP。除了共同的蛋白质亚基形成的核心颗粒之外，每一种 snRNP 还有各自的辅助蛋白质因子。

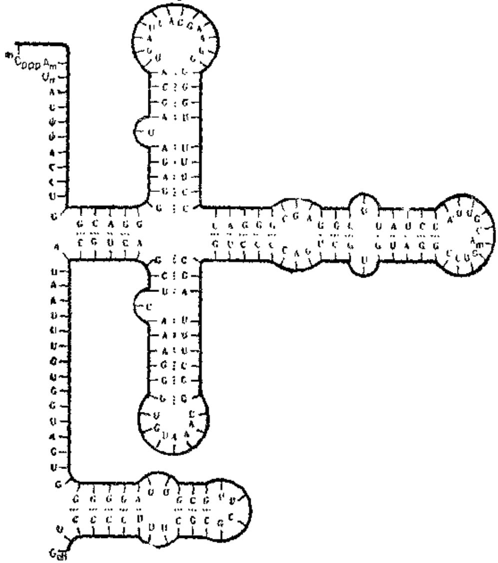


图 6-49 人类 U1 snRNA 的二级结构，其中 5′ 末端处于单链状态，其他物种亦可能有类似的序列和结构

在各种 snRNA 中，只有 U1 snRNA的 5′ 端序列能够与 mRNA 前体的两个拼接点序列互补。U1 snRNA 已发现于哺乳动物，鸟类，昆虫的细胞中。人类 U1snRNP 中除了 RNA 之外，还含有 8 个蛋白质亚基。人类 U1RNA 的可能的二级结构示于图 6-49。其中 5′ 末端的 11 个核苷酸呈单链状态，因此能够与拼接点序列互补。snRNA 与拼接点配对情况的模式图示于图 6-50。上图表示 snRNA 与内元相互结合的总体结构；下图为一个典型的内元序列与 snRNA 的碱基配对状况，其中 U1 snRNA 的 CCU 序列的精确配对情况还很难预言，因为它与两个拼接点的 AGG 都能互补。图 6-50 所示为最好的结合状况。对于大多数内元来说，其左拼接点有 4～6 碱基配对，其右拼接点只有两个碱基配对。在体外试验中，U1 snRNP 只能与左拼接点结合，目前尚无与右拼接点结合的证据。与拼接点结合是整个 U1 snRNP 的性质，提纯了的 snRNA 不能与拼接点结合。我们知道，这些少量的特异性的碱基配对必须依赖于特定的环境，就像 tRNA 的反密码子的三个碱基与密码子配对依赖于核糖体一样。U1 snRNA 与其它蛋白质可能构成一个复杂的拼接体(spliceosome)，U1 snRNA 只有在这样的一个复合体中才能识别拼接点序列。

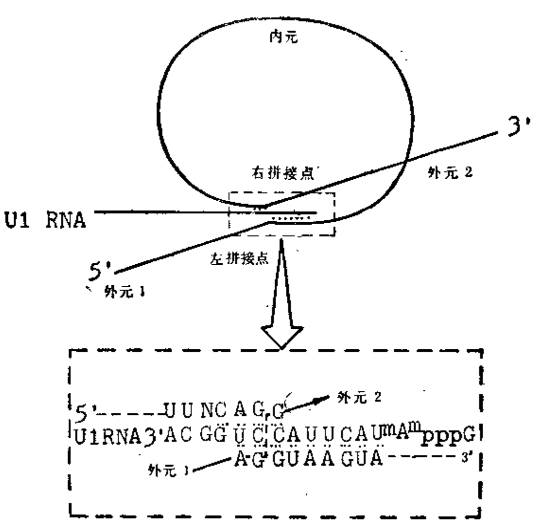



图 6-50　U1 RNA 与内元的左拼接点配对，也同右拼接点配对，以实现两个拼接点的断裂与再连接

关于 U1 snRNP 参与核 mRNA 前体拼接的证据很多。最初的一个间接证据就是一种自身免疫疾病——红斑狼疮病人体内有一种抗体，其主要作用对象是 snRNP。HeLa 细胞受腺病毒感染后，腺病毒 RNA 能够在细胞中拼接。如果将这种细胞核分离出来与抗 snRNP 的抗体保温，则拼接反应停止。用补骨脂素(psoralen)处理核蛋白，然后再用紫外线处理，则可以发现有 U1 snRNA 释放出来。因为补骨脂素只能交联双链核酸分子。也就是说，snRNA 已经与 mRNA 配对结合，分离提取后再用紫外线解交联，snRNA 才能释放出来。

在真核生物核 mRNA 前体的拼接体中，除了 U1 snRNP 之外，还发现其他种类的 snRNP：U2 snRNP，U5 snRNP，U4/U6snRNP。迄今为止，人们并没有发现这些 snRNP 具有酶的活性。它们都是与 mRNA 前体共同组成一个拼接体。U1 snRNP 通过其 RNA 识别 5′ 端拼接点序列并与之结合；U2 snRNP 通过其 RNA 识别并结合分枝点；U5 snRNP 通过其 70KDa 辅助蛋白质结合于 3′ 拼接点，其具体位置在 3′ 拼接点与分枝点之间富含嘧啶的序列。这个 70KDa 蛋白质是 Sm 蛋白质的一个新成员，为 U5 snRNP 所特有。在其它已发现的 Sm 成员中，A 蛋白质，C 蛋白质和 68KDa 蛋白为 U1 snRNP 所特有；B 蛋白(28KDa)，B′ 蛋白(29KDa)、D 蛋白(16KDa)、E 蛋白(13KDa)等均为各 snRNP 的通用成分，其中 D 蛋白直接与各种 U RNA 分子上的所谓 Sm 接合位点发生作用。这些 U snRNP 除了与 mRNA 前体发生相互作用之外，还可相互之间发生作用。U4/U6 snRNP 可能并不与 mRNA 前体结合，而是充当拼接体的结构成分并可能承担各种 UsnRNP 之间的"通讯"工作。

真核生物的 RNA 拼接体大约有 50～60S。用 HeLa 细胞核提取液和 mRNA 前体可以进行拼接体的组装。首先形成一个 30～35S 的复合体，这一过程是需要 ATP 的；然后才形成一个 50～60S 的有功能的拼接体。其具体过程还不清楚，据推测可能是先由 U5 snRNP 识别 3′

拼接点和分枝点，然后由 U2 snRNP 识别分枝点，最后由 U1 snRNP 识别 5′ 拼接点。除了 snRNP 之外，可能还需要 hnRNP。在 mRNA 前体上，有三段重要的序列，即 5′ 拼接点的 GU，分枝点，3′ 拼接点的 AG 及其前面的 11 个富含嘧啶的序列。分枝点位于内元 3′ 末端上游方向大约 30～40 个碱基处，其一致序列为 $P_yXP_yTP_u\overset{*}{A}P_y$。其拼接过程与第 Ⅱ 类内元一样，也是由这个保守 $\overset{*}{A}$ 对 5′ 拼接点发动亲核进攻，使内元与前一个外元断开，内元 5′ 端的 G与保守$\overset{*}{A}$通过5′,2′ 酯键形成套索结构中间体(包括内元及其下游外元)。然后与 U1snRNP 相连接的切下来的外元的 3′ 末端进攻内元的 3′ 拼接点，切除呈套索状的内元，连接两个外元。实验表明，当 5′ 拼接点切断后，U1 snRNP 就连接于 3′ 拼接点，这对于后一步反应至关重要。U1 snRNP 是通过结合 U5 snRNP 而与 3′ 拼接点相连结还是 U1 snRNP 取代了 U5 snRNP 而与 3′ 拼接点结合？现在还不清楚，看来有可能是后一种情况，其中可能存在着一种别构效应。

## 第九节　转录后处理(四)　不连续转录和反式拼接[113～119]

1982 年，van der Ploeg 等人在研究锥虫 (*Trypahosome*)的可变表面糖蛋白(VSG)基因及其 mRNA 时发现，在 mRNA 5′ 端有 35 个碱基是其基因中所没有的。这就表明，这两段 RNA 是分别转录的，然后才通过某种机制连接到一起去的。后来发现，锥虫的全部 mRNA 都含有这个 35 碱基序列；而这些 mRNA 基因中全都没有这样的序列。这样的 35 个碱基的序列，就称为小外元(miniexon)。后来人们发现这种小外元的基因呈连续重复的形式排列在锥虫的染色体上，大约有 200 个这样的重复单位。每个基因转录产生 135 个核苷酸的产物，前35 个为外元，后 100 个为内元，存在着一个真核生物典型的 5′ 拼接点。外元序列有 60% 的保守性，而内元序列除了拼接点之外，无保守性(图 6-51)。

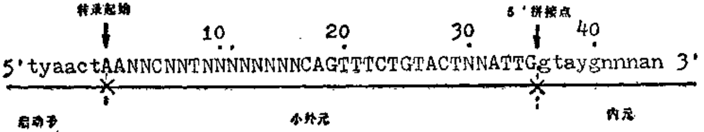



图 6-51　小外元基因结构

小外元有 60% 的保守序列；内元的前 9 个核苷酸也有 60% 的保守性；转录起始点前 6 个核苷酸亦高度保守。内元只有 5′ 拼接点，并具备真核生物的保守序列gt。小外元用大写字母表示，其余序列用小写字母。

这个小外元到底是通过什么方式与 mRNA 的编码序列相连接？先后有三种假说出现。一种是引物假说，即小外元基因产物作为引物，然后共价延伸产生 *VSG* 基因转录物，再通过顺式拼接除去内元。在当时，发现 *VSG* 基因的转录是对鹅膏蕈碱不敏感的，这表明转录 *VSG* 基因的不是 RNA 聚合酶 Ⅱ，而是一种特殊的 RNA 聚合酶，它需要小外元基因产物(包括其内元转录出的序列)作为引物来转录 *VSG* 基因。后来查明，其他蛋白质基因的转录全是 RNA 聚合酶 Ⅱ 进行的，而 RNA 聚合酶 Ⅱ 是不需要引物的。现在已经知道，*VSG* 基因是 RNA 聚合酶 Ⅰ 转录的，而 RNA 聚合酶 Ⅰ 在锥虫体内转录的各种 rRNA 都不需要小外元产物作为引物。因而这种假说早已被否定。第二种假说是这两种基因转录产物由 RNA 连接酶首

先连接成一条 RNA 分子，然后再通过顺式拼接除去内元。第三种假说是，这两种独立的基因产物，不产生共线性的中间产物，而是直接将两个外元拼接在一起。这种拼接方式与已发现的 RNA 拼接方式(顺式)不同，称为反式拼接。上述三种假说，从 RNA 拼接的角度来说，可以归结为两类，前两种为顺式拼接，最后一种为反式拼接。

1986 年，W.J.Murphy 和 R.E.Sutton 等人差不多同时证明了锥虫小外元与蛋白质基因的外元的拼接方式是反式拼接。他们发现拼接产物是 Y 型结构，而不是套索结构。他们将拼接产物用 HeLa 细胞核中的去分枝酶(特异地作用于 2′,5′ 磷酸二酯键)处理，结果分离到 100nt 的小外元基因产物中的内元序列。如果是顺式拼接的话，只能得一个很大的内元序列(图 6-52)。

反式拼接的机制，人们未设想它与顺式拼接有什么本质的不同。它仍然是一连串的磷酸酯转移反应。在锥虫这一类低等真核生物中已发现许多反式拼接现象。这种反式拼接，亦可能需要一个拼接体来促成拼接反应的完成，催化作用仍然是 RNA 本身所具备的性质。

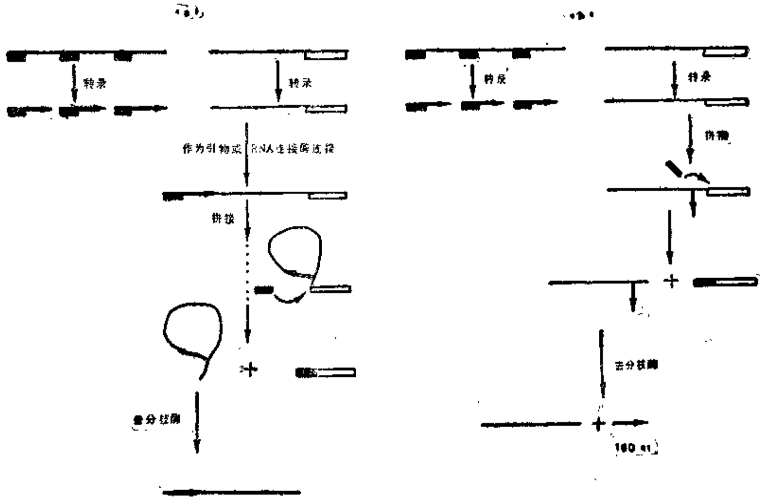



图 6-52 小外元序列与蛋白质基因的mRNA两种不同拼接方式及其结果
(a) 顺式拼接；(b) 反式拼接

奇怪的是，在锥虫的蛋白质基因中，没有一个编码序列是被内元所打断的。因此在锥虫体内只有反式拼接，而不存在顺式拼接。现在还不知道为什么这两种拼接方式不能共存，很可能在锥虫体内只存在着一种适应反式拼接的拼接体。关于反式拼接的生物学意义还不清楚，可能与锥虫体内 mRNA 的戴帽子作用以及基因表达的调控有关。关于反式拼接，最近在高等真核生物中，如烟草的叶绿体中亦有发现。

# 第十节 逆 转 录[120~123]

严格说来，逆转录并不属于转录的范畴。不仅逆转录的过程和产物与转录完全相反，而且介导这一过程的逆转录酶与 RNA 聚合酶截然不同。从本质上来说，逆转录酶是属于 DNA 聚合酶，其前体物为 dNTP(dATP、dGTP、dCTP、dTTP)，产物为 DNA 链；必须有引物存在的情况下才能起始聚合；$Zn^{++}$ 为这类金属蛋白所不可缺少的成分。从生物学意义上来说，转录的目的在于制造基因的产物(RNA 或者蛋白质)，而还原病毒逆转录的直接目的在于基因组的重组。因此，关于逆转录的过程和意义主要放在第十一章(遗传重组)中讲授。本节仅扼要地叙述逆转录酶的结构及其酶学特性。安排本节内容的目的仅在于表明 DNA 与 RNA 之间的关联是通过 RNA 聚合酶和逆转录酶而实现的。由于逆转录酶和逆转录作用的发现，才使中心法则得到补充和完善(图 6-53)。

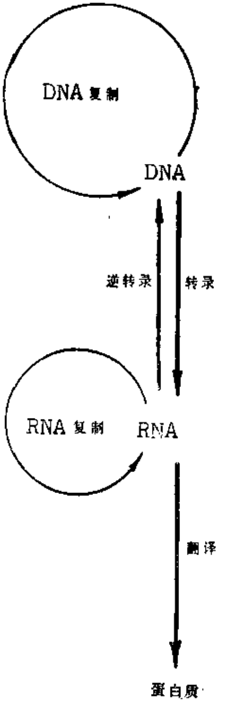



图 6-53 逆转录过程的发现完善了中心法则。

逆转录酶是 1960 年由美国威斯康星大学 Temin 等人首先发现的。他们研究鸟类的路斯氏肉瘤病毒时发现了一种依赖于 RNA 的 DNA 聚合酶，也就是人们发现的第一个逆转录酶。后来，人们陆续发现了各种高等真核生物的 RNA 肿瘤病毒都有逆转录酶。这些 RNA 病毒统称为还原病毒。在果蝇等真核生物中，还发现了活动的散布遗传成分 (mobile d spersed genetic elements，简写mdg)，其结构与还原病毒非常相似(详见第十一章)。所有这些还原病毒都有三个"基因"，一个为gag，编码病毒种族特异性的抗原(group-specific antigen)；一个是pol，编码逆转录酶；一个是 env，编码病毒外膜蛋白。这三个"基因"只是指其编码序列，而实际上通过不同的转录后和翻译后处理可以产生多种不同的蛋白质。

鸟类的 RNA 病毒 AMV(avian myeloblastosis virus)，ASV(avian sarcoma virus)的逆转录酶具有两个亚基。其中$\alpha$亚基分子量为 63KDa，$\beta$ 亚基为 94KDa。$\alpha$ 亚基是 $\beta$ 亚基经蛋白质水解作用后产生的，而 $\beta$ 亚基的前体可能是 $P_r180^{gag-pol}$。鸟类的逆转录酶有三种分子形式，一种是 $\alpha$，一种是 $\alpha\beta$、一种是 $\beta_2$，其中 $\alpha\beta$ 是主要形式，对病毒 70S RNA 的亲和力也最强。哺乳类还原病毒的逆转录酶一般为单亚基结构。这些病毒可以分为 B、C、D 三种类型。B 型以小鼠乳瘤病毒(MMTV)为代表，其逆转录酶分子量为 100KDa；C 型以鼠类白血病病毒(MuLV)为代表，其逆转录酶分于量为 80KDa；D 型以 Mason Pfizer 氏猴病毒(MPMV)为代表，其逆转录酶分子量大约为 90KDa。

逆转录酶是一类非常复杂的酶，在其简单的一条肽链上具有多种酶活性。在组成形式为 $\alpha\beta$ 的逆转录酶中，酶活性由 $\alpha$ 亚基提供。主要的酶活性有 DNA 聚合酶活性、RNAaseH 活性、DNA 内切酶活性、DNA 旋转酶活性、螺旋酶活性以及 tRNA 结合活性等。下面分别叙述

这些酶活性。

1. DNA聚合酶活性 逆转录酶不但具有RNA指导的DNA聚合酶活性，而且具有DNA指导的DNA聚合酶活性。因此，逆转录酶能够以RNA或DNA为模板，在引物(RNA、DNA均可)存在的情况下合成DNA。尽管两种DNA聚合酶活性的热失活速度不同，但是人们认为这两种活性是同一的。逆转录酶与所有的DNA聚合酶一样需要$Zn^{++}$，需要一个带有3′-OH并与模板以氢键结合的事先形成的引物。DNA链的合成方向为5′→3′，即通过在引物的3′-OH与5′-三磷酸脱氧核苷前体之间形成3′,5′磷酸二酯键而使DNA链进行共价延伸。当初人们还以为逆转录是不需要引物的，其实是病毒RNA分子已经结合了tRNA分子(在ASV RNA分子上结合了$tRNA^{Trp}$)。因此还原病毒的粒子中有四条RNA链通过氢键相互连接起来。两个tRNA链分别与两条病毒基因组RNA链以反平行方式联结；而两条病毒基因组RNA链则以平行方式联结，对于这种联结方式，人们至今还缺乏详细的了解。

人们利用不同的模板——引物体系可以区分还原病毒的逆转录酶与细胞DNA聚合酶。逆转录酶能利用$(Cm)_n\cdot(dG)_{15}$-进行有效的DNA合成而不能利用$(dT)_n\cdot(A)_{15}$-和$(dA)_n(dT)_{15}$-；而细胞DNA聚合酶恰恰相反。然而，并非仅有逆转录酶才能以RNA为模板合成DNA，人们发现*E.coli* DNA聚合酶I能同样有效地以RNA为模板拷贝出DNA链。

逆转录酶不具备3′→5′和5′→3′外切酶活性。真核生物的DNA聚合酶α、β、γ虽然也缺少3′→5′外切酶活性，但可能由于具有3′→5′外切酶活性的DNA聚合酶δ的配合，真核生物的DNA复制与*E.coli*具有同样水平的保真度甚至更高一些。然而，逆转录酶拷贝出的DNA保真度则要低得多，错配碱基的出现几率要比真核生物或*E.coli*高1～3个数量级，这就是各种RNA肿瘤病毒具有很高的自发突变频率的原因。

逆转录酶的作用方式是进行性的(processive)还是非进行性的，仍有争论。一般认为，逆转录酶αβ是进行性的，α是非进行性的，而哺乳类还原病毒的逆转录酶的作用方式仍未得出结论。

2. RNAaseH活性 还原病毒要将自己的单链RNA分子转变为双链DNA分子才能整合到寄主染色体中。要完成这样一个过程，就必需有降解RNA的酶活性(这一复杂过程将在第十一章中讲述)。因此，所有的逆转录酶都具有RNAaseH的活性。然而逆转录酶的RNAaseH活性一般是属于外切酶性质(5′→3′和3′→5′)，而不具备内切酶性质，因此作用较慢。逆转录酶αβ的RNAaseH活性是进行性的，而其他的逆转录酶的RNAaseH的作用方式则可能是随机的。

有许多证据表明，RNAaseH活性与DNA聚合酶活性位于同一条肽链上，但位于不同的区域。一般说来，DNA聚合酶活性比较容易受到热和化学试剂的作用而失活，而RNAaseH则有较大的抗性。用蛋白水解酶将AMV的逆转录酶α降解成24KDa的片断或将MuLV的逆转录酶降解成30KDa的片断，结果都失去了DNA聚合酶活性而只剩下RNAaseH活性。

3. DNA内切酶活性 所有的逆转录酶都表现出一定的DNA内切酶活性，这种酶活性具有位点特异性。当有$Mn^{++}$存在时，则失去了位点特异性，而成为对超螺旋更为敏感的切刻酶活性。DNA内切酶活性可能与重组功能有关。

4. DNA旋转酶活性 据报道，ASV病毒粒子含有DNA旋转酶活性，能够使I型DNA变为I°型。这方面的细节及其生理功能都还不清楚。

5. 螺旋酶活性 据报道，AMV的逆转录酶αβ或α都具有螺旋酶活性。解旋结果可以

用具有单链特异性的核酸酶 S1 来进行检验。这种活性只作用于 DNA·DNA 双链和 DNA·RNA 双链，而不能作用于 RNA·RNA 双链。同样有证据表明，这一酶活性与 DNA 聚合酶活性、RNAaseH 活性位于同一条肽链的不同区域。

6. tRNA 结合活性 所有的逆转录酶都具有与其引物 tRNA 结合的能力。在 ASV 和 AMV 中为 $tRNA^{Trp}$，在 MuLV 中为 $tRNA^{Pro}$，在果蝇的 mdg1 中为 $tRNA_2^{Arg}$。这一能力显然是这些还原病毒或返座元(retroposon)起始逆转录的重要保证。

## 提　要

在 RNA 聚合酶的催化之下，以 DNA 为模板合成 RNA 的整个过程，包括起始、延伸、终止等步骤，叫做转录。转录是基因表达的第一步，也是最关键的一步，也是基因调控的主要步骤。在双链 DNA 中，不作为转录模板的链称为有义链或编码链，而作转录模板的链称为反义链或模板链。

RNA 聚合酶与 DNA 聚合酶不同，它能起始一条新链的合成而不需要引物。大肠杆菌的 RNA 聚合酶"全酶"通常指 $\alpha_2\beta\beta'\sigma$ 这样五个亚基。后来发现的 NusA 也可以看作是它的亚基。$\sigma$ 亚基是全酶识别启动子的要素；$\beta$ 亚基能与底物结合并催化磷酸二酯键的形成；$\beta'$ 亚基可能与有义链结合；$\alpha$ 亚基可能参与全酶与启动子的牢固结合，并与 RNA 聚合酶通过时局部 DNA 螺旋的改变有关。真核生物有三种 RNA 聚合酶，分别转录 rRNA 基因、蛋白质基因和 5S rRNA、tRNA 基因。此外，细胞器还有自己的 RNA 聚合酶。

原核生物中，功能相关的基因通常前后相连成为操纵元，由一个共同的控制区进行转录的控制。控制区通常包括操作子和启动子两个部分。启动子有两个重要的序列，即 −10 序列和 −35序列。许多分解代谢操纵元，在其启动子中还有 cAMP-CAP 结合位点。真核生物的三种 RNA 聚合酶有着各自的启动子类型。RNA 聚合酶 Ⅰ 只有一种启动子类型；RNA 聚合酶 Ⅲ 通常为基因内启动子(下游启动子)；RNA 聚合酶 Ⅱ 的启动子最为复杂，有帽子位点和 TATA 框等近启动子成分，而上游启动子成分则因基因而异，常见的有 CAAT 框，GC 框等。真核生物的基因还常含有增强子。增强子成分与上游启动子成分一样，均通过其蛋白质结合因子才能促进转录。

原核生物 RNA 聚合酶的 $\sigma$ 亚基能发现 −35序列，全酶与-35序列结合后形成封闭启动子复合物。然后全酶以某种尚未知晓的机制转移到 −10序列并诱导了 −10 序列的局部解链，形成开放性启动子复合物。$\beta$ 亚基上的起始位点和延长位点被相应的核苷酸充满并形成第一个磷酸二酯键，从而构成开放的三元复合物。然后 $\sigma$ 亚基脱落，致使三元复合物中核心酶与DNA 的亲和力下降，以便转录的顺利进行。RNA 链延长阶段的最大特点是速度不恒定，有时会发生延宕。真核生物转录的起始机制还不清楚。

转录的终止需要终止子结构。终止子是在 RNA 水平上发生作用的。终止子转录成 RNA 后形成柄-噜噗结构，与 $\beta$ 亚基的空间结构相嵌合。不依赖于 $\rho$ 因子的终止子在柄部富含 G·C 碱基对，而且紧跟着柄—噜噗结构的是富含 A/U 序列，特别是一连串的 U。而依赖于 $\rho$ 因子的终止子则相反。$\rho$ 因子的活性形式为六聚体，能结合于新生 RNA 链并利用水解NTP 所释放的能量从 5′ 向 3′ 方向移动。RNA 聚合酶在终止子处的延宕，给予 $\rho$ 因子以追赶的机会。$\rho$ 因子通过与 $\beta$ 亚基的作用，促进转录的终止。原核生物的转录和翻译的耦联使 $\rho$ 因子结合到 RNA 上的机会较少。依赖于 $\rho$ 因子的终止子前的无义突变则造成后续基因无法转录。

这就是无义突变的极性效应。RNA聚合酶进行转录时另一个终止因子就是NusA亚基。NusA不能与全酶结合,而能与核心酶结合,而且这种结合是非序列特异性的。λ噬菌体的抗终止蛋白N和Q结合于*nut*和*qut*位点,与一个分子的NusA,NusB,NusE以及已经结合于RNA聚合酶的一个NusA构成延伸控制颗粒,从而实现抗终止作用。真核生物三类RNA聚合酶的转录终止子结构及其终止机制还不清楚,可能都需要富含A/T序列。特别是RNA聚合酶Ⅱ的转录物的多聚腺苷酸化使得探索其转录终止机制变得更为困难。

几乎全部的真核mRNA的5′端都具有帽子结构,大部分的mRNA的3′端具有多聚A尾巴。帽子有三种类型,5′末端的$N^7$-甲基鸟苷酸以5′-5′三磷酸二酯键与帽子位点(即转录的第一个核苷酸)相连,使得mRNA并无自由的5′末端。多聚腺苷酸化常常发生在转录终止之前,由识别AAUAAA(以及GUGUGUG)的特异的酶以及poly(A)聚合酶完成。这两种酶可能与其它蛋白质成分和RNA成分组成RNP颗粒才能完成多聚腺苷酸化。

原核生物的三种rRNA基因与tRNA基因混杂排列成7个rrn操纵元。转录产生的30S的多基因RNA产物,然后由RNAaseⅢ切除先导序列,拖尾序列和基因间序列。真核生物的5.8S、18S、28S三种rRNA组成一个转录元,转录产生47S的前体RNA分子。这种RNA分子上有大约110个甲基化位点(核糖2′-O上),这是处理酶系的识别标志。tRNA转录产物的处理酶系十分复杂,其中值得注意的是RNAaseP,它由375核苷酸的$M_1$ RNA和20 KDa的$C_5$蛋白质构成。

不连续基因转录产生的原初转录物中含有相应的内元序列,这些内元序列必须去除之后才能成为成熟的有功能的RNA。根据内元拼接点的序列特征,可以分为tRNA内元,第Ⅰ类内元,第Ⅱ类内元和核mRNA内元四大类型。tRNA的内元由特异的酶切除,然后由RNA连接酶将两个tRNA半分子连接起来。第Ⅰ类内元具有保守的拼接点序列U↓……G↓,能形成中部核心结构的保守的5′-P-Q-R-S-3′序列以及能与两个拼接点序列配对的内部引导序列。通过鸟核苷或鸟苷酸的亲核进攻,发起一连串的磷酸酯转移反应,最后去除了内元,将外元连接起来。有些内元能够与外元联合编码RNA成熟酶,这个成熟酶是去除该内元所不可缺少的。第Ⅱ类内元除了↓GUGCG……AU↓这一拼接点特征外,在3′拼接点附近还有$P_yP_uP_yP_yT\overset{*}{A}P_y$,其中$\overset{*}{A}$总是处于二级结构之外。正是这个$\overset{*}{A}$的2′羟基发动亲核进攻,进行一连串的转酯反应,切下套索状的内元并将外元连接起来。核mRNA内元的拼接模式与第Ⅱ类内元十分相似,只是需要极为复杂的拼接体。拼接体是由U系snRNA和一系列的蛋白质构成。核mRNA内元有保守的5′拼接点↓GU和3′拼接点AG↓以及分枝点$P_yXP_yTP_u\overset{*}{A}P_y$。同样,由$\overset{*}{A}$的2′羟基发动亲核进攻,切除套索状的内元并将外元连接起来。RNA催化作用的发现,对于研究生命的进化给予极其巨大的推动。

锥虫的蛋白质基因虽然含有内元,但并非不连续。蛋白质基因前面的内元含有3′拼接点,小外元基因后面的内元含有5′拼接点。这两类基因分别转录成RNA,两种RNA分子的外元再拼接起来形成成熟的mRNA,两部分内元形成Y型产物。这种拼接方式就是反式拼接。

从DNA拷贝出RNA的过程称为转录;与此相反,从RNA拷贝出DNA的过程称为逆转录。逆转录酶对于研究遗传重组和生命的进化有着重要的意义。逆转录酶和逆转录作用的发现补充和完善了中心法则。

# 参 考 文 献


1. Crick, F. H. C. (1958) "On Protein Synthesis." *Symp. Soc. Exp. Biol.* 12:548～555.
2. Tisseires, A. et al. (1959) "Ribonucleoprotein Particles From *Escherichia coli*." *J. Mol. Biol.* 1:221.
3. Watson, J. D. (1963) "Involvement of RNA in Synthesis of Proteins." *Science* 140: 17～26.
4. Baltimore, D. (1971) "Expression of Animal Virus Genomes" *Bacteriol. Revs.* 35:235～241
5. Hurwitz, J. et al. (1960) "The Enzymatic Incorporation of Ribonucleotides into Polyribonucleotides and the Effect of DNA." *Biochem. Biophys. Res. Comm.* 3:15～19.
6. Stevens, A. (1960) "Incorporation of the Adenine Ribonucleotide into RNA by Cell Fractions from *E. coli B*." *Biochem. Biophys. Res. Comm.* 3: 92～96.
7. Weiss, S. B. (1960) "Enzymatic Incorporation of Ribonucleotide Triphosphates into the Interpolynucleotide Linkages of Ribonucleic acid." *Proc. Nat. Acad. Sci.* 46:1020～1030.
8. Taylor, K. et al. (1967) "Asymmatric Distribution of the Transcribing Regions on the Complementary Strands of Coliphage λ DNA." *Proc. Nat. Acad. Sci.* 57:1618～1625.
9. Burton, Z.F. et al. (1983) "The Operon that Encodes the Sigma Subunits of RNA Polymerase also Encodes Ribosomal Protein S21 and DNA Primase in *E.coli K12*." *Cell* 32:335～349.
10. Travers, A.A. and Burgess, R.R (1969) "Cyclic Reuse of the RNA Polymerase Sigma Factor." *Nature* 222: 537～540.
11. Hanna, M. and Meares, C. (1983) "Topography of Transcription, Path of the Leading end of Nascent RNA through the *Escherichia coli* Transcription Complex." *Proc. Nat. Acad. Sci.* 80:4238～4242.
12. Sommerville, J. (1984) "RNA Polymerase I Promoters and Transcription Factors." *Nature* 310:189～190.
13. Nevins, J.R. (1983) "The Pathway of Eukaryotic mRNA Formation." *Ann. Rev. Biochem.* 52:441～466
14. Ciliberto, G. et al. (1983) "Transcription by RNA Polymerase Ⅲ." *Curr. Top. Dev. Biol.* 18:59～88.
15. Schmitz, A. and Galas, D. (1979) "The Interaction of RNA Polymerase and lac Repressor with the *lac* Control Region." *Nucleic Acid Res.* 6:111～137.
16. Siebenlist, U. et al. (1980) "*E.coli* RNA Polymerase Interacts Homologously with Two Different Promoters." *Cell* 20:269～281.
17. Youderian, P. et al. (1982) "Sequence Determinants of Promoter Activity." *Cell* 30:843～853.
18. Hawley, D. K. and McClure, W. R. (1983) "Compilation and Analysis of *Escherichia coli* Promoter DNA Sequences." *Nucleic Acid Res.* 11:2237～2255.
19. Beckwith, J.R. and Zipser, D., eds (1970) *The Lactose Operon.* Cold Spring Harbor Laboratory.
20. Jacob, F. and Monod, J. (1961) "Genetic Regulatory Mechanisms in the Synthesis of Proteins." *J. Mol. Biol.* 3:318～356.
21. Ptashne, M. and Hopkins, N. (1968) "The Operators Controlled by the Phage Repressor." *Proc. Nat. Acad. Sci.* 60:1282～1287.
22. Zubay, G. et al. (1970) "Mechanism of Activation of Catabolite-Sensitive Genes, A Posistive Control System." *Proc. Nat. Acad. Sci.* 66:104～110.
23. DeCrombrugghe, B. et al. (1984) "Cyclic AMP Receptor Protein: Role in Transcription Activation." *Science* 224:831～838.
24. McKnight, S.L. and Kingsbury, R. (1982) "Transcriptional Control Signals of a Eukaryotic Protein-coding Gene." *Science* 217: 316～324.
25. Shenk, T. (1981) "Transcriptional Control Regions, Nucleotide Sequence Requirements for Initiation by RNA Polymerase Ⅱ and Ⅲ." *Curr. Top. Microbiol. Immunol.* 93:25～46.
26. Fire, A. et al. (1984) "Interactions Between RNA Polymerase Ⅱ, Factors, and Template Leading to Accurate Transcription." *J. Biol. Chem.* 259:2509～2516.
27. Dynan, W. S. and Tjian, R. (1985) "Control of Eukaryotic Messenger RNA Synthesis by Sequence-specific DNA Binding Proteins." *Nature* 316:774～777.
28. Brown, D. D. (1984) "The Role of Stable Complexes That Repress and Activate Eucaryotic Genes." *Cell* 37: 359～365.
29. Lassar, A.B. et al. (1983) "Trancription of Class Ⅲ Genes, Formation of Preinitation Complexes." *Science* 222:740～748.

30. Melnikova, A. F. et al. (1978) "A study of Unwinding of DNA and Shielding of the DNA Grooves by RNA Polymerase by Using Methylation with Dimethylsulphate." *Eur. J. Biochem.* **84**: 301〜309.

31. Gamper, H. and Hearst, J. (1982) "A Topological Model for Transcription Based on Unwinding Angle Analysis of *E. coli*. RNA Polymerase Binary, Initiation, and Ternary Complexes." *Cell* **29**:81〜90.

32. Horwitz, R. J. et al. (1987) "An Elongation Control Particle Containing the N Gene Transcriptional Antitermination Protein of Bacteriophage $\lambda$." *Cell*, **51**:631〜641.

33. McClure, W. R. and Hoopes, B. C. (1985) "Regulation of Transcription Initiation in *Escherichia coli*." in *Sequence Specificity In Transcription And Trrnslation*, ed. Calendar, R and Gold, L., New York, Alan R. Liss, pp. 11〜20.

34. Zheng, X-M. et al. (1987) "A General Transcription Factor Forms a Stable Comlex with RNA Polymerase B (II)." *Cell* **50**:361〜368.

35. Roberts, J. W. (1969) "Termination Factor for RNA Synthesis." *Nature* **224**:1168〜1174.

36. Lowery-Goldhammer, C. and Richardson, I. P. (1974) "An RNA-dependent Nucleoside Triphosphate Phosphohydrolase (ATPase) Associated with Rho Termination Factor." *Proc. Nat. Acad. Sci.* **71**: 2003〜2007.

37. Ryan, T. and Chamberlin, M. (1983) "Transcription Analyses with Heteroduplex Trp Attenuator Templates Indicate that the Transcript Stem and Loop Structure Serves as The Termination Signal." *J. Biol. Chem.* **258**: 4690〜4693.

38. Platt, T. and Bear, D. (1983) "The Role of RNA Polymerase, Rho Factor, and Ribosomes in Transcription Termination." in *Gene Function In Procaryotes* ed. Beckwith, J. et al, Cold Spring Harbor Laboratory, pp. 123〜162.

39. Platt, T. (1986) "Transcription Termination and the Regulation of Gene Expression." *Ann. Rev. Biochem.* **55**: 339〜372.

40. Fisher, R. and Yanofsky, C. (1983) "Mutations of the $\beta$ Subunit of RNA Polymerase Alter Both Trancription Pausing and Transcription Termination in the Trp Operon Leader Region in Vitro." *J. Biol. Chem.* **258**: 8146〜8150.

41. Greenblatt, J. and Li, J. (1981) "The NusA Gene Protein of *Escherichia coli*: Its Identification and a Demonstration that It Interacts with the Gene N Transcription Antitermination Protein of Bacteriophage Lambda." *J. Mol. Biol.* **147**:11〜23.

42. Greenblatt, J. and Li, J. (1981) "Interaction of the Sigma Factor and the Nus A Gene Protein of *E. coli* with

52. Proudfoot, N. J. (1989)"How RNA Polymerase Ⅱ Terminates Transcription in Higher Eukaryotes" *Trends Biochem. Sci.* 14: 105～110.
53. Perry, R.P. and Kelly, D. E.(1976)"Kinetics of the Formation of 5 Terminal Caps in mRNA." *Cell* 8:433～442.
54. Kozak, M.(1978)"How Do Eucaryotic Ribosomes Select Initiation Regions in Messenger RNA?" *Cell* 15:1109～1123.
55. Shuman, S. and Hurwitz, J. (1982) "Capping Enzyme." in *The Enzymes*, ed. Boyer, P. D., New York, Academic Press, 15: 246～265.
56. Shatkin, A. J. (1985)"mRNA Cap Binding Proteins: Essential Factors for Initiating Translation." *Cell* 40: 223～224.
57. McLaughlin, C. S. et al.(1973)"Polyadenylic Acid Sequences in Yeast Messenger Ribonucleic Acid." *J. Biol. Chem.* 248: 1466～1471.
58. Edmonds, M. (1982) "Poly(A) Adding Enzymes."in *The Enzymes*, ed. Boyer, P. D., New York: Academic, Press15: 218～244.
59. Birnstiel, M. L. et al.(1985) "Transcription Termination and 3 Processing: The End Is in Site!" *Cell* 41: 349～359.
60. Hashimoto, C. and Steitz, J. A (1986) "A Small Nuclear Ribonucleoprotein Associates with the AAUAAA Polyadenylation Signal in Vitro."*Cell* 45:581～591.
61. Kessler, M. M. et al.(1986)"Requirement of AAUAAA and Adjacent Downstream Sequences for SV40 Early Polyadenylation."*Nucleic Acid Res.* 14:4939～4952.
62. Takagaki et al. (1988) "Separation and Characterization of a Poly(A) Polymerase and a Cleavage/Specificity Factor Required for Pre-mRNA Polyadenylation"*Cell* 52:731～742
63. Brimacombe, R.(1984)"Conservation of Structure in Ribosomal RNA."*Trends in Biochem. Sci.*9:273～277.
64. Gegenheimer, P. and Apirion, D. (1981)"Processing of Procaryotic Ribonuclic Acid."*Microbiol Rev.* 45: 502～541.
65. Grummt, I. et al.(1986)"A Transcription Terminator Located Upstream of the Mouse rDNA Initiation Site Affects rRNA Synthesis."*Cell* 47:901～911.
66. Labhart, P. and Reeder, R. H. (1987)"Ribosomal Precursor 3 End Formation Requires a Conserved Element Upstream of the Promoter."*Cell* 50:51～57.
67. Dunn, J.J.(1982)"RibonucleaseⅢ"*The Enzymes*, ed. Boyer, P.D., 15:485～499.
68. Kole, R. and Altman, S.(1982)"tRNA Processing Enzymes from *Escherichia coli*."in *The Enzymes*, ed. Boyer P.D., 15:470～483.
69. Deutscher, M.P.(1982)"tRNA Nucleotidyltransferase."in *The Enzymes*, ed. Boyer, P.D., 15:183～215.
70. Gold, H. A. and Altman, S.(1986)"Reconstitution of RNAaseP Activity Using Inactive Subunits from. E. coli and HeLa Cells."*Cell* 44:243～249.
71. Guerrier-Takada, C. et al.(1983)"The RNA Moiety of RibonucleaseP Is the Catalytic Subunit of the Enzyme." *Cell* 35: 849～857.
72. Abelson, J. (1979) "RNA Processing and the Intervening Sequence Problem."*Ann. Rev. Biochem.* 48:1035～1069.
73. Davies, R. W. et al (1982)"Making Ends Meet: A Model for RNA Splicing in Fungal Mitochondria." *Nature* 300: 719～724.
74. Michel, F. and Dujon, B.(1983)"Conservation of RNA Secondary Structures in Two Intron Families Including Mitochondrial-, Chloroplast-, and Nuclear-encoded Members." *EMBO J.* 2: 33～38.
75. Peebles, C. L. et al.(1986)"A Self-splicing RNA Excises an Intron Lariat."*Cell* 44:213～223.
76. van der Veen, R. et al.(1986)"Excised Group Ⅱ Intron in Yeast Mitochondria Are Lariats and Can Be Formed by Self-splicing in Vitro." *Cell* 44:225～234.
77. Cech, T.R. (1983)"RNA splicing: Three Themes with Variations."*Cell* 34:713～716.
78. Greer, C.L. et al.(1983)"Mechanism of Action of a Yeast RNA Ligase in tRNA Splicing."*Cell* 32:537～546.
79. Laski, F.A. et al.(1983)"Characterization of tRNA Precursor Splicing in Mammalian Extracts."*J. Biol. Chem.* 258:11974～11980.
80. Baldi, M. I. et al. (1986)"Binding and Cleavage of Pre-tRNA by the Xenopus Splicing Endonuclease: Two Separable Steps of the Intron Excision Reaction."*Cell* 47:965～971.

81. Cech, T. R. et al. (1983) "Secondary Structure of the *Tetrahymena* Ribosomal RNA Intervening Sequence: Structural Homology with Fungal Mitochondrial Intervening Sequences." *Proc. Nat. Acad. Sci.* **80**:3903～3907.
82. Cech, T. R. (1985) "Self-splicing RNA: Implications for Evolution." *Int. Rev. Cytol.* **93**:3～22.
83. Cech, T. R. (1986) "RNA as an Enzyme." *Sci. Amer.* **255** 64～75.
84. Cech, T. R. and Bass, B. L. (1986) "Biological Catalysis by RNA." *Ann. Rev. Biochem.* **55**:599～629.
85. Garriga, G. et al. (1986) "Mechanism of Recognition of the 5'Splice Site in Self Splicing Group I Introns." *Nature* **322**: 86～89.
86. Szostak, J. W. (1986) "Enzymatic Activity of the Conserved Core of a Group I Self-splicing Intron." *Nature* **322**:83～86.
87. Cech, T. R. (1986) "A Model for the RNA-catalyzed Replication of RNA." *Proc. Nat. Acad. Sci.* **83**: 4360～4363.
88. Dinter-Gottlieb, G. (1986) "Viroids and Virusoids Are Related to Group I Introns." *Proc. Nat. Acad. Sci.* **83**: 6 250～6 254.
89. Epstein, L. M. and Gall, J. G. (1987) "Self-cleaving Transcripts of Satellite DNA from the Newt." *Cell* **48**:535～543.
90. Gilbert, W. (1986) "The RNA world" *Nature* **319**:618.
91. Cech, T. R. (1989) "Ribozyme Self-Replication?" *Nature* **339**:507～508
92. Doudna, J. A. and Szostak, J. W. (1989) "RNA-Catalysed Synthesis of Complementary-Strand RNA." *Nature* **339**:519～522.
93. Chu, F. K. et al. (1986) "Characterization of the Intron in the Phage T4 Thymidylate Synthase Gene and Evidence for Its Self-Excision from the Primary Transcript." *Cell* **45**:157～166.
94. Lazowska, J. et al. (1980) "Sequence of Introns and Flanking Exons in Wild-type and Box-3 Mutants of Cytochrome b Reveals an Interlaced Splicing Protein Coded by An Intron." *Cell* **22**: 333～348.
95. Borst, P. and Grivell, L. A. (1981) "One Gene s Intron Is Another Gene s Exon." *Nature* **289**:439～440
96. Garriga, G. and Lambowitz, A. L. (1984) "RNA Splicing in Neurospora Mitochondria: Self-splicing of a Mitochondial Intron in Vitro." *Cell* **39**:631～641.
97. Schmelzer, C. and Schweyen, R. J. (1986) "Self-splicing of Group II Introns in Vitro: Mapping of the Branch Point and Mutational Inhibition of Lariat Formation." *Cell* **46**:557～565.
98. Jacquier, A. and Michel, F. (1987) "Multiple Exon-binding Sites in Class II Self-splicing Introns." *Cell* **50**: 17～29.
99. Breathnach, R. and Chambon, P. (1981) "Organization and Expression of Eucaryotic Split Genes Coding for Proteins." *Ann. Rev. Biochem.* **50**:349～383.
100. Mount, S. M (1982) "A Catalogue of Splice Junction Sequences." *Nucleic Acid Res.* **10**:459～472.
101. Treisman, R. et al. (1983) "Specific Transcription and RNA Splicing Defects in Five Cloned β-thalassaemic Genes." *Nature* **302**:591～596.
102. Mattag, I. W. (1984) "snRNA: From Gene Architecture to RNA Processing." *Trends Biochem. Sci.* **9**:435～437.
103. Lerner, M. R. et al. (1980) "Are snRNPs Involved in Splicing?" *Nature* **283**:220～224.
104. Mount, S. M. et al. (1983) "The U1 Small Nuclear RNA-protein Complex Selectively Binds a 5' Splice Site in Vitro." *Cell* **33**: 509～518.
105. Ruskin, B. et al. (1984) "Excision of an Intact Intron as a Novel Lariat Structure During Pre-mRNA Splicing in Vitro." *Cell* **38**:317～331.
106. Brody, E. and Abelson, J. (1985) "The Spliceosome': Yeast Premessenger RNA Associates with a 40s Complex in a Splicing Dependent Reaction." *Science* **228**:963～967.
107. Berget, S. M. and Robberson, B. L. (1986) "U1, U2, and U4/U6 Small Nuclear Ribonucleoproteins Are Required for in Vitro Splicing but Not Polyadenylation." *Cell* **46**:691～696.
108. Gerke, V. and Steitz, J. A. (1986) "A Protein Associated with Small Nuclear Ribonucleoprotein Particles Recognizes the 3' Splice Site of Premessenger RNA". *Cell* **47**:973～984.
109. Parry, H. D. et al. (1989) "Snurpogenesis: The Transcription and Assembly of U snRNP Components." *Trends Biochem. Sci.* **14**:15～19.
110. Konarska, M. M. and Sharp, P. A. (1987) "Interactions Between Small Nuclear Ribonucleoprotein Particles in Formation of Spliceosomes." *Cell* **49**:763～774.
111. Reed, R. et al. (1988) "Purification and Visualization of Native Spliceosomes." *Cell* **53**:949～961.

112. Cech, T. R. (1986) "The Generality of Self-spicing RNA: Relationship to Nuclear mRNA Splicing." *Cell* 44: 207～210.
113. Heinrichs, V. et al. (1990)"U1-Specific Protein C Needed for Efficient Complex Formation of U1 snRNP with a 5' Splice Site."*Science* 247:69～72.
114. Borst, P.(1986)"Discontinuous Trascription and Antigenic Variation in *Trypanosomes*." *Ann. Rev. Biochem.* 55: 701～732.
115. van der Ploeg, L. (1986)"Discontinuous Transcription and Splicing in *Trypanosomes*."*Cell* 47:479～480.
116. Murphy, W.J. et al.(1986)"Identification of a Novel Y Branch Structure in *Trypanosome* mRNA Processing: Evidence for Trans Splicing."*Cell* 47:517～525.
117. Sutton, R.E. et al(1986)"Evidence for Trans Splicing in *Trypanosomes*."*Cell* 47:527～535.
118. Koller, B. et al. (1987)"Evidence for in Vivo Trans Splicing of Pre-mRNAs in Tobacco Chloroplasts." *Cell* 48: 111～119.
119. Sharp, P.A. (1987)"Trans-splicing: Variation on a Familiar Theme?"*Cell* 50:147～148.
120. Temin, H.M. and Mizutani, S.(1970)"Viral RNA-dependent DNA Polymerase."*Nature* 226: 1211～1213.
121. Baltimore, D.(1970)"Viral RNA-dependent DNA Polymerase."*Nature* 226: 1209～1211.
122. Verma, I.M.(1982)"Reverse Transcriptase" in *The Enzymes*, ed. Boyer, P. D. Academic Press. 14:87～103.
123. Varmus, H. E. (1989)"Reverse Transcription in Bacteria."*Cell* 56:721～724.

# 第七章 翻 译

对于终产物为 RNA 的基因，只要进行转录并进行转录后的处理，就完成了基因表达的全过程；而对于终产物是蛋白质的基因，还必须将 mRNA 翻译成蛋白质。mRNA 的核苷酸顺序与蛋白质的氨基酸顺序之间在结构上并没有直接的相应关系，二者也不发生直接的相互作用。在这两种不同的遗传语言之间，必须通过译员才能互相沟通。扮演这种译员角色的就是各种

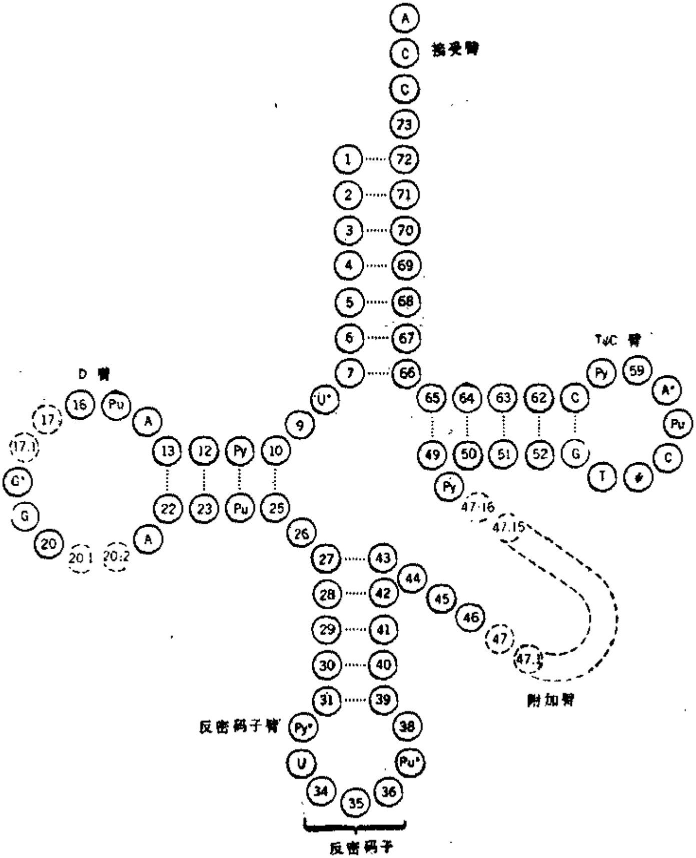



图 7-1 tRNA 的二级结构

实线圆圈中的数字代表从 5′ 端数起的碱基的顺序号，实际上可为任何一个碱基；而字母则代表保守或半保守的碱基，虚线圆圈则表示该碱基并不存在于所有的 tRNA 分子中

tRNA 分子。如果没有 tRNA 的存在，也就无所谓密码子了。因此密码子的意义并不是单独由 mRNA 决定的，而是由 mRNA 和携带氨基酸的 tRNA 相互作用所决定的。然而，tRNA 分子上的反密码子与其相应的氨基酸之间并无结构上的相关性，而且 tRNA 分子本身并不能催化相应氨基酸的荷载，而是需要相应的氨酰基-tRNA 合成酶来催化。那么，氨酰基-tRNA 合成酶又是如何识别它的两个底物的呢？对于氨基酸来说，该酶无疑是识别它们的侧链基团；而对于 tRNA 分子来说，氨酰基-tRNA 合成酶又是识别哪个基团呢？原来，每个 tRNA 分子上有一个特异的小区以供其氨酰基-tRNA 合成酶所识别，这就是 tRNA 分子上的副密码子。荷载相应氨基酸的 tRNA 分子并不能独立地与 mRNA 发生作用，因为单靠密码子与反密码子三个碱基对之间的氢键联结是无法保证配对精确性的，更无法形成肽键并促成肽链的起始，延伸和终止。这就需要核糖体和有关因子的存在。在遗传信息流动的各个过程中，参与翻译过程的生物大分子数目最多。因此，翻译过程是一个极为复杂的过程。

## 第一节　tRNA 和遗传密码[1~20]

### 一、tRNA 的结构

tRNA 的结构相当保守，各种物种的 tRNA 均含有 70～80 个碱基，tRNA 均具有三叶草形的二级结构(图 7-1)和 L 状的三级结构(图 7-2)。从二级结构上看，tRNA 分子可以分为五个主要的臂：携带氨基酸的接受臂；TψC 臂；反密码子臂；双氢尿嘧啶(DHU)臂；附加臂。前四个臂序列保守性强，而附加臂大小不一。附加臂的大小常常反映了整个 tRNA 分子的大小。根据附加臂的大小，可将 tRNA 分为两类：大约 3/4 的 tRNA 只含有 3～5 个碱基的附加臂，称为第 I 类 tRNA；其余的 tRNA 常含有 13～21 个碱基的附加臂，称为第 II 类 tRNA。

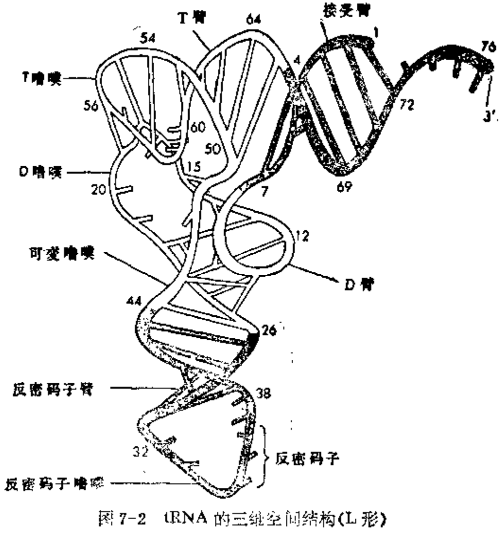



图 7-2　tRNA 的三维空间结构(L 形)

将 tRNA 分子的二级结构画成三叶草状，只是为了方便醒目，并不代表真实的空间构型。tRNA 分子的实际结构要紧密得多。tRNA 结晶的 X 光衍射结果表明，所有的 tRNA 分子均为 L 形状。根据三叶草模型所推测的各个臂确实存在于空间结构中，而且采取人们所预期的 A 构象。在接受臂上也确实存在着 G·U 配对现象。三维结构主要也是靠碱基之间氢键力维持的。形成三级结构氢键的碱基在二级结构中大多没有形成氢键。许多保守的或半保守的碱基参与了三级结构氢键的形成。这就意味着所有的 tRNA 分子可能具有同样形状的空间结构。在生理条件下，tRNA 分子可能具有更大的灵活性和可变性。在 L 形结构中，接受臂和反密码子臂分别位于两条双螺旋的顶端，这样使得 tRNA 所携带的氨酰基靠近核糖体大亚基的肽基转移酶位点，而反密码子能够与小亚基上的 mRNA 密码子配对。TψC 嘧啶和 DHU 嘧啶位于 L 的两臂交界处，可能有利于控制 tRNA 三维结构。在 tRNA 的三维结构中，除了氢键作用力之外，相邻碱基堆集力也是维持其空间结构的重要因素(图 7-3)。

## 二、密码子与反密码子

DNA 或 mRNA 的四种碱基共组成 64 个三联体密码子(codons)，其中 61 个密码子编码常规的 20 种氨基酸，3 个密码子为肽链终止密码，又称为无义密码子。无义密码子并非总是无义，除了无义抑制 tRNA 能够识别它们以外，现已发现携带稀有氨基酸的 tRNA 也能够识别终止密码子，而且是这些稀有氨基酸渗入肽链的正常途径。这些稀有氨基酸如磷酸丝氨酸，硒半胱氨酸等(见下一小节)。由于有 61 个密码子分配给 20 种氨基酸，因此出现了由多种密码子(最多为 6 种)编码一个氨基酸的现象，这称为密码子的简并性(degeneracy)。由 6 个密码子编码一个氨基酸如丝氨酸时，是否需要 6 种丝氨酸 tRNA 来识别它们呢？不是的。一种 tRNA 分子常常能够识别一种以上的同一种氨基酸的密码子，这是因为 tRNA 分子上的反密码子与密码子的配对具有摇摆性(wobble)，也就是说一种反密码子能够与不同的

表 7-1 按照摇摆假说反密码子与密码子的碱基配对

| 反密码子碱基 | 密码子碱基 |
|---|---|
| G | C或U |
| C | G |
| A | U |
| U | A或G |
| I | U,C或A |

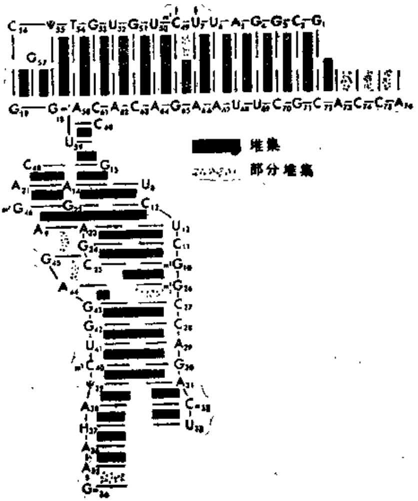



图 7-3 酵母苯丙氨酸tRNA分子中的碱基堆集作用。深色表示完全堆集，浅色表示部分堆集。较粗的线条表示核苷酸的嘌呤环或嘧啶环。不参与堆集作用的碱基(第16，17，20-47 碱基)则省略。

密码子发生碱基配对。配对的摇摆性完全是由 tRNA 反密码子嘌呤的空间结构所决定的。如上所述，在 tRNA 分子的空间结构中，相邻碱基存在着碱基堆集力。然而，反密码子 5′ 端的一个碱基却处在一连串堆集碱基的末端，它受的堆集力较小，因此有较大的自由度(图7-3)。而且，反密码子 5′ 端碱基常为修饰过的碱基，很少有 U，而几乎没有 A；A 被修饰为 I（次黄嘌呤），它可以和 U，C，A配对。加上 RNA 中常见的 G·U 配对，因此配对情况就复杂得多(表 7-1)。根据反密码子与密码子的摇摆配对，人们推测至少存在着三种丝氨酸 tRNA，后来也证实了这种推测。这样，携带氨基酸相同而反密码子不同的一组 tRNA 称为同功tRNA(isoaccepting tRNAs)。一组同功 tRNA 由同一种氨酰基-tRNA 合成酶催化而携带氨基酸。

有时，具有相同反密码子（携带相同的氨基酸）的 tRNA 也有好几种，它们的结构有很大的差异，例如具有同样的反密码子 3′-AUG-5′ 的两种酪氨酸 tRNA。这些 tRNA 是由不同的基因编码的。其中有的在细胞内含量很高，称为多数 tRNA(major tRNA)，有的含量很低，称为少数 tRNA(minor tRNA)。二者浓度之差可以高达 11 倍，这是由于各自启动子强度不同而引起的。少数 tRNA 的功能还不完全清楚，现在已经知道的就是许多抑制 tRNA 基因都是由少数 tRNA 基因突变而成(见第五章)。

### 三、副密码子与氨酰基 tRNA 的合成

氨酰基-tRNA 是由氨酰基-tRNA 合成酶催化产生的。这种酶需要三种底物即氨基酸，tRNA 和 ATP。ATP 提供了活化氨基酸所需要的能量。氨基酸首先活化成氨酰基腺苷酸(这一中间产物与酶相结合)，然后再与 tRNA 生成氨酰基 -tRNA。其反应式如下：

$$AA + ATP \rightleftharpoons AA - AMP + ppi$$

$$AA - AMP + tRNA \rightleftharpoons AA - tRNA + AMP$$

总反应：$$AA + tRNA + ATP \rightleftharpoons AA - tRNA + AMP + ppi$$

细胞内有 20 种氨基酸，有 40种以上的 tRNA，氨酰基 tRNA 合成酶是如何选择正确的氨基酸和 tRNA 的呢？人们至今尚未弄清其详细作用机制。按照一般原理，酶和底物的正确结合是由二者相嵌的几何形状所决定的。只有适合的氨基酸和适合的 tRNA 进入合成酶的相应位点，才能合成正确的氨酰基-tRNA。现在已经知道合成酶与 L 形 tRNA 的内侧面结合，结合点包括接受臂，DHU 臂和反密码子臂。在这几个位点上的 tRNA 突变则妨碍正确的氨基酸接载。

乍看起来，反密码子似乎应该与氨基酸的正确荷载有关，对于某些 tRNA 也确实如此。然而对于大多数 tRNA 来说，情况并非如此。人们早就知道，来源于少数 tRNA 的抑制 tRNA，尽管它们的反密码子已经改变(识别无义密码子)，但它们携带氨基酸的性质并没有改变。1986 年，Normanly 等人将亮氨酸 tRNA 分子中置换了 12 个核苷酸(反密码子未动)，则变成了携带丝氨酸的 tRNA。1988 年，侯雅明和 Schimmel 做了十分出色的试验：将丙氨酸抑制 tRNA ($tRNA^{Ala/CUA}$)分子突变了许多位点，组成了28种突变体(14 种在接受臂上)。结果发现，许多突变体都不改变接载丙氨酸的性质，而只有改变 G3:U70 这一碱基对的突变体才表现出明显的突变效应(不能抑制 trpA 基因中第 234 个 amber 无义突变)。另外两种 amber 抑制 tRNA ($tRNA^{Cys/CUA}$，$tRNA^{Phe/CUA}$) 虽然对于 trpA 基因中的无义突变能够抑制，但都不能恢复 $TrpA^+$ 表现型，因为半胱氨酸或苯丙氨酸的插入都不能形成有活性的 α 亚基。尽管这两个抑制 tRNA 分别有 38 个和 31 个核苷酸不同于 $tRNA^{Ala/CUA}$，但是如果将它们的 C3:G70 改变

为 G3:U70，均能为丙氨酸 tRNA 合成酶所识别。这就说明，G 3:U 70 是丙氨酸 tRNA 分子决定其本质(携带丙氨酸)的主要因素。tRNA 分子上决定其携带氨基酸的区域被称为副密码子(paracodon)。

由上所述，可以看出，一种氨酰基 tRNA 合成酶可以识别一组同功 tRNA (最多达 6 个)，这说明它们具有共同特征。例如丙氨酸 tRNA 合成酶识别 G3:U70，在已发现的三种丙氨酸 tRNA($tRNA^{Ala/CUA}$, $tRNA^{Ala/GGC}$, $tRNA^{Ala/UGC}$)都具有G3:U70副密码子。但还没有充足的证据说明其他氨酰基 tRNA 合成酶在其同功 tRNA组中也是识别相同的副密码子。其次，副密码子并没有固定的位置，亦可能并不止一个碱基对。例如上述的 $tRNA^{Cys/CUA}$和$tRNA^{Phe/CUA}$都具有 C3:G70，显然它们的副密码子不在 C3:G70 或者不完全在 C3:G70，因为它们携带两种不同的氨基酸。第三，尽管副密子不是单独与氨基酸发生作用，但是副密码子可能与氨基酸的侧链基团有某种相应性。tRNA 分子对错误接载的氨基酸具有校对功能，可能就是来源于这种相应性。副密码子可能是从立体化学上决定相应氨基酸的"原始副密码子"(protopar-acodon)进化而来。人们早就推测，tRNA 分子起源于能够携带氨基酸的寡聚核苷酸。这种识别氨基酸的特异性在某种程度上保留下来，而反密码子是在 tRNA 分子进化过程中后来才出现的。这就是说，副密码子比密码子和反密码子更为古老。

关于氨酰基-tRNA 合成酶和 tRNA 参与校对的分子细节还不清楚。该系统可以通过多级的校对功能排除错误的接载。其中主要有动力学校对 (Kinetic proofreading)，构象校对(conformational proofreading)和化学校对(chemical proofreading)。动力学校对是基于合成酶对于错误氨基酸形状的识别，即使生成氨酰基腺苷酸形式，也会将它水解。构象校对是在氨酰基腺苷酸生成以后，由于 tRNA 的进入引起酶构象的改变，使得已生成的氨酰基腺苷酸不符合进一步连接成氨酰基-tRNA 的要求，从而水解之。化学校对是指错误的氨基酸已接载到 tRNA上之后，被合成酶的 tRNA 结合位点所发现，然后水解之(图 7-4)。虽然说不少人倾向于最后一步机制，但都缺少令人信服的证据。

### 四、tRNA 的丰富度与密码子的使用频率

大多数氨基酸是由一个以上的密码子所编码。这个事实则提出了一个问题：这种编码同一种氨基酸的一组密码子的使用频率是否都相同？人们推测，在富含 A+T 的基因组中，密码子的第三位上 U 和 A 出现的几率较多。例如单链 DNA 噬菌体 $\phi$X174 基因组由24%的A，22%的 C，23% 的 G 和 31% 的 T 组成，其序列分析结果证实了在 $\phi$X174 的密码子的第三位上 U 的出现几率最多。

在 A+T 含量和 G+C 含量没有显著差异的基因组中，密码子的使用频率也不是平均的。有些密码子使用率很高，有些则几乎不使用。现在，将大肠杆菌中含量最丰富的核糖体蛋白质基因中密码子使用频率列于表 7-2。在高等真核生物中，密码子的使用也不是随机的，其偏爱的密码子与不喜欢的密码子当然并不一定与大肠杆菌相同，只不过差异比大肠杆菌要小一些(表 7-3)。

细致的分析表明，密码子的使用频率与细胞内 tRNA 含量(即 tRNA的丰富度)呈正相关，特别是对于那些表达量高的蛋白质基因更是如此 (如核糖体蛋白质，RecA 蛋白质)。在真核生物中，情况也十分类似，相应于主要同功 tRNA 的密码子的使用频率几乎总是最高的。酵母中含量丰富的乙醇脱氢酶同工酶 I 和 3-磷酸甘油醛脱氢酶便是最好的例子。在这两个基因中，

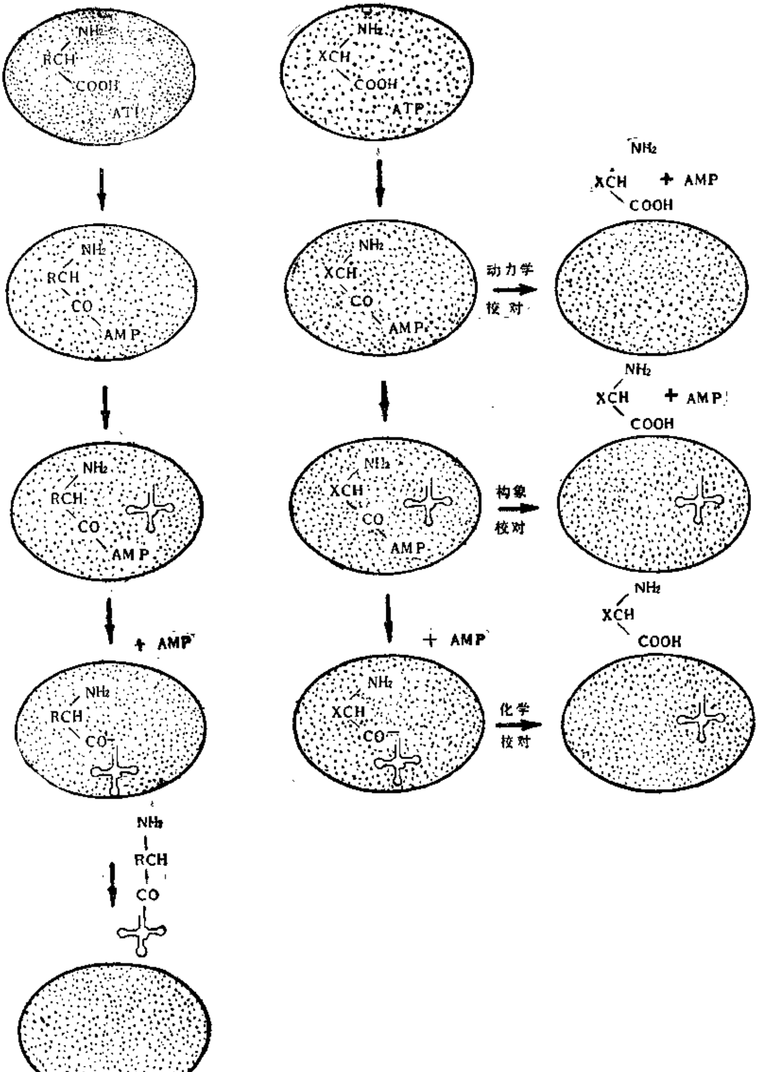



图 7-4　在氨基酰-tRNA 合成酶作用的不同阶段对接载的氨基酸进行鉴别——多级校对

RCH($NH_2$)COOH 表示正确的氨基酸

XCH($NH_2$)COOH 表示错误的氨基酸

**表 7-2 大肠杆菌某些核糖体蛋白质中密码子使用频率**

（表中所示为 1 209 个密码子中出现次数）

| 密码子的第 1 位 | 密码子的第 2 位 U | C | A | G | 密码子的第 3 位 |
|---|---|---|---|---|---|
| U | 10 UUU Phe<br>23 UUC Phe<br>1 UUA Leu<br>2 UUG Leu | 18 UCU Ser<br>18 UCC Ser<br>1 UCA Ser<br>1 UCG Ser | 3 UAU Tyr<br>13 UAC Tyr<br>UAA *Stop*<br>UAG *Stop* | 1 UGU Cys<br>6 UGC Cys<br>UGA *Stop*<br>3 UGG Trp | U<br>C<br>A<br>G |
| C | 4 CUU Leu<br>3 CUC Leu<br>0 CUA Leu<br>79 CUG Leu | 3 CCU Pro<br>0 CCC Pro<br>4 CCA Pro<br>36 CCG Pro | 3 CAU His<br>15 CAC His<br>9 CAA Gln<br>33 CAG Gln | 48 CGU Arg<br>26 CGC Arg<br>0 CGA Arg<br>0 CGG Arg | U<br>C<br>A<br>G |
| A | 13 AUU Ile<br>51 AUC Ile<br>0 AUA Ile<br>30 AUG Met | 36 ACU Thr<br>26 ACC Thr<br>3 ACA Thr<br>0 ACG Thr | 3 AAU Asn<br>42 AAC Asn<br>90 AAA Lys<br>24 AAG Lys | 1 AGU Ser<br>12 AGC Ser<br>1 AGA Arg<br>0 AGG Arg | U<br>C<br>A<br>G |
| G | 54 GUU Val<br>6 GUC Val<br>40 GUA Val<br>16 GUG Val | 93 GCU Ala<br>10 GCC Ala<br>45 GCA Ala<br>28 GCG Ala | 17 GAU Asp<br>45 GAC Asp<br>61 GAA Glu<br>16 GAG Glu | 49 GGU Gly<br>34 GGC Gly<br>0 GGA Gly<br>0 GGG Gly | U<br>C<br>A<br>G |

**表 7-3 哺乳动物基因中密码子使用频率**

（表中所示为 2 244 个密码子中出现的次数）

| 密码子的第 1 位 | 密码子的第 2 位 U | C | A | G | 密码子的第 3 位 |
|---|---|---|---|---|---|
| U | 13 UUU Phe<br>28 UUC Phe<br>2 UUA Leu<br>9 UUG Leu | 16 UCU Ser<br>18 UCC Ser<br>9 UCA Ser<br>2 UCG Ser | 10 UAU Tyr<br>23 UAC Tyr<br>UAA *Stop*<br>UAG *Stop* | 10 UGU Cys<br>13 UGC Cys<br>UGA *Stop*<br>12 UGG Trp | U<br>C<br>A<br>G |
| C | 9 CUU Leu<br>27 CUC Leu<br>7 CUA Leu<br>47 CUG Leu | 14 CCU Pro<br>17 CCC Pro<br>10 CCA Pro<br>5 CCG Pro | 10 CAU His<br>21 CAC His<br>10 CAA Gln<br>28 CAG Gln | 8 CGU Arg<br>11 CGC Arg<br>4 CGA Arg<br>5 CGG Arg | U<br>C<br>A<br>G |
| A | 11 AUU Ile<br>24 AUC Ile<br>4 AUA Ile<br>16 AUG Met | 15 ACU Thr<br>28 ACC Thr<br>11 ACA Thr<br>6 ACG Thr | 8 AAU Asn<br>28 AAC Asn<br>19 AAA Lys<br>49 AAG Lys | 12 AGU Ser<br>21 AGC Ser<br>8 AGA Arg<br>10 AGG Arg | U<br>C<br>A<br>G |
| G | 9 GUU Val<br>21 GUC Val<br>6 GUA Val<br>33 GUG Val | 28 GCU Ala<br>38 GCC Ala<br>14 GCA Ala<br>6 GCG Ala | 16 GAU Asp<br>24 GAC Asp<br>21 GAA Glu<br>34 GAG Glu | 22 GGU Gly<br>32 GGC Gly<br>16 GGA Gly<br>11 GGG Gly | U<br>C<br>A<br>G |

96% 以上的氨基酸只用 25 种密码子。这样，细胞能够以更快的速度合成需要量大的蛋白质；而需要量少的蛋白质基因中具有较多的与含量少的 tRNA 相应的密码子，用以控制该蛋白质的合成速度。这也是原核生物和真核生物基因表达调控的共同战略之一。大肠杆菌中 tRNA 的丰富度与其使用频率的相关性示于图 7-5。

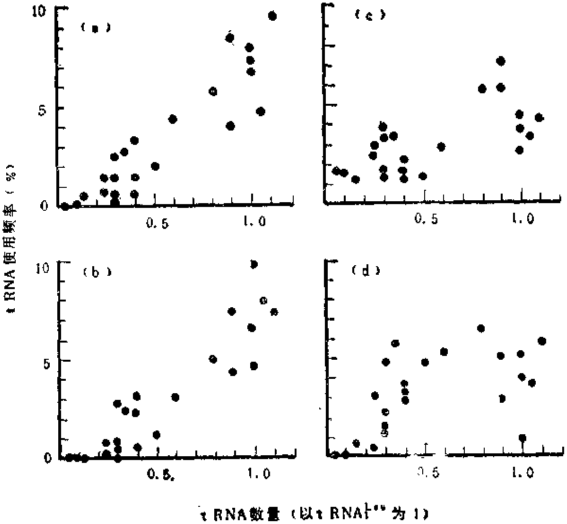



图 7-5 大肠杆菌26种tRNA的丰富度(以$tRNA_1^{Leu}$的含量作为1.0)与某些基因或一组基因中的使用频率的相关性。注意在表达量高的基因recA基因(a)和核糖体蛋白质基因(b)中，二者呈明显的正相关；而在表达量低的基因β-内酰胺酶基因(c)和在表达量有高有低的一组φX174基因(d)中，虽然也有某种程度的正相关，但表现更大程度的离散分布。

在同一种tRNA分子所识别的几种同义密码子(最多为3种)中，其使用频率也有明显的差异。这可能主要是由反密码子与密码子的作用强度所决定的。适中的作用最利于翻译过程的迅速进行；弱配对作用可能使荷载的tRNA分子进入A需要花费更多的时间；强配对作用可能使转肽以后对P位上已卸载的tRNA分子的逐出要消耗更多的时间。利用这种理论可以解释如下的事实(大肠杆菌中)：(1) 密码子GGG(Gly)，CCC(Pro)，AUA(Ile)的使用率都几乎为0；(2) 密码子前两位由A和U组成的同义密码子中，第三位为C的密码子的使用频率要高于U或A，即：UUC>UUU，UAC>UAU，AUC>AUU，AAC>AAU。此外，反密码子的摇摆碱基如果是修饰的U，则与A配对的机会多于G；如果摇摆碱基是I，与U和C配对几率多于A。

### 五、密码的通用性、线粒体密码的特殊性和密码的进化

用同聚核苷酸作为mRNA，在无细胞体系中进行的翻译试验表明，poly U总是刺激苯丙氨酸的渗入，polyC总是刺激脯氨酸的渗入，polyA总是刺激赖氨酸的渗入，poly G总是刺激甘氨酸的渗入。这种无细胞抽提液不管是来自细菌还是来自动物组织，都产生同样的结果。现在大量的事实已经证明了生命世界从低等到高等，都使用几乎完全相同的一套密码。也就是说，遗传密码子在很长的进化时期中保持不变。因此，表7-2和7-3所列的密码是一种通用密码。

然而，出乎人们预料的是，真核生物线粒体的密码子有若干处不同于通用密码子(以人

类线粒体为例）：

(1) UGA 不是终止密码子，而是色氨酸的密码子。这符合摇摆规律。

(2) AGA，AGG不是精氨酸的密码子，而是终止密码子。这样，加上通用密码中的UAA和UAG，线粒体共有四个终止密码子。

(3) 内部甲硫氨酸密码子有两个，即AUG和AUA；起始甲硫氨酸密码子有四个，即AUN；

(4) 在酵母线粒体中，除上述三点与通用密码子不同之外，密码子 CUA 不是亮氨酸的密码子，而是苏氨酸的密码子。

总之，线粒体的密码子表排列得比较整齐，各种氨基酸的密码子以及起始和终止密码子的数目或是 2，或是4，或是 6。密码子数目为 2 的有苯丙氨酸、异亮氨酸、酪氨酸、组氨酸、谷酰胺、天门冬酰胺、赖氨酸、天门冬氨酸、谷氨酸、半胱氨酸、色氨酸、甲硫氨酸(但起始甲硫氨酸密码子有 4 个)。数目为 4 的有缬氨酸、脯氨酸、苏氨酸、丙氨酸、精氨酸、甘氨酸和终止密码。有 6 个密码子的只有亮氨酸和丝氨酸。人类线粒体的遗传密码列于表 7-4。

**表 7-4 人类线粒体的遗传密码**

| 密码子的第 1 位 | 密码子的第 2 位: U | C | A | G | 密码子的第 3 位 |
|---|---|---|---|---|---|
| U | UUU, UUC Phe (GAA)<br>UUA, UUG Leu (UAA) | UCU, UCC, UCA, UCG Ser (UGA) | UAU, UAC Tyr (GUA)<br>UAA *Stop*<br>UAG *Stop* | UGU, UGC Cys (GCA)<br>[UGA], UGG Trp (UCA) | U<br>C<br>A<br>G |
| C | CUU, CUC, CUA, CUG Leu (UAG) | CCU, CCC, CCA, CCG Pro (UGG) | CAU, CAC His (GUG)<br>CAA, CAG Gln (UUG) | CGU, CGC, CGA, CGG Arg (UCG) | U<br>C<br>A<br>G |
| A | AUU, AUC Ile (GAU)<br>[AUA], AUG Met (CAU) | ACU, ACC, ACA, ACG Thr (UGU) | AAU, AAC Asn (GUU)<br>AAA, AAG Lys (UUU) | AGU, AGC Ser (GCU)<br>[AGA] *Stop*<br>[AGG] *Stop* | U<br>C<br>A<br>G |
| G | GUU, GUC, GUA, GUG Val (UAC) | GCU, GCC, GCA, GCG Ala (UGC) | GAU, GAC Asp (GUC)<br>GAA, GAG Glu (UUC) | GGU, GGC, GGA, GGG Gly (UCC) | U<br>C<br>A<br>G |

注：与通用密码有差别的部分用方框表示。每个方括号由同一种 tRNA 所识别的一组密码子；圆括号内的序列为tRAN的反密码子(5′→3′)；甲硫氨酸 tRNA 反密码子的第一个核苷酸为C，它与 AUA 的非寻常配对机制还不清楚。

线粒体密码的整齐划一性质是与其 tRNA数目少相一致的。根据摇摆配对学说，使用通用密码子的有机体至少得具备 32 种 tRNA 分子，而人类线粒体只有 22 种 tRNA 分子，只有具有 6种密码子的亮氨酸和丝氨酸才各有两个 tRNA 基因。当一个 tRNA 分子识别四个一组的密码子时，其反密码子的第一位(摇摆位置)均为 U，能与密码子第三位上任何一个碱基配对，其机制还不清楚。当密码子的第三位为嘧啶和嘌呤而分别编码两种不同的氨基酸时，则相应 tRNA 的反密码子 5′ 端核苷酸分别为 G 和经过修饰的 U。经过修饰的 U 只能识别密码子第三位的嘌呤。线粒体 tRNA 还有许多与细胞 tRNA 不同的地方，如细胞 tRNA 分子中保守的 GTψCRA 序列在大多数线粒体 tRNA 分子中都不存在。T 噜噗和 D 噜噗上还有不少差别，

丝氨酸 tRNA 则完全缺乏细胞 tRNA 的 D 臂。这些差别暗示着线粒体 tRNA 在三维折叠方面和与其核糖体作用方面均不同于细胞 tRNA。

线粒体的较为简单而整齐的密码子安排到底是代表着比"通用"密码更为古老的遗传密码还是"通用"密码的流线化形式呢？现在还无法得出结论。其中涉及到的不仅是密码子的起源问题，而且有线粒体的起源问题，即使假定目前多数人赞同的内共生假说是正确的，还有内共生作用发生在生命进化的哪个时期的问题。下面，编者列出一些有利于线粒体密码的古老性假说的论据。

首先，从终止密码子的变化来看，人类线粒体只编码 13 个蛋白质，似乎没有必要将终止密码子由三个变为四个。在大肠杆菌中，甘氨酸抑制 tRNA ($tRNA^{Gly/UCU}$) 对于精氨酸密码子 AGA 的错义抑制有时高达 25%。如果 AGA 是常用的典型精氨酸密码子，则不可能演化出这种致命的错义抑制 tRNA。甘氨酸抑制 tRNA本来可能不是错义抑制 tRNA，而可能是无义抑制 tRNA，其功能是抑制 AGA 和 AGG这两个终止密码子。这样，一方面能够矫正无义突变，另一方面又能有效地利用 mRNA 模板，产生长短不同的两种蛋白质。这就像通用密码中 UGA 这一终止密码子的渗漏性一样。

大肠杆菌的色氨酸 tRNA 在正常情况下就能以一定的几率识别通用密码中的终止密码子 UGA。这说明 UGA 本来很可能就是色氨酸密码子。只有在释放因子 RF2 演化产生以后，UGA 才转变为终止密码子。然而 UGA是一种渗漏终止密码子，给有机体带来基因表达的更大灵活性。例如大肠杆菌噬菌体 Qβ 的外壳蛋白(131 个氨基酸)基因以 UGA 为终止密码子。翻译时有 95% 的几率在此释放肽链，产生主要的外壳蛋白。而有 5% 的几率则通读而产生327个氨基酸组成的 A1 蛋白，最后在双重终止密码子 UAGUAA 处释放肽链。A1 蛋白是 Qβ 侵染寄主所不可缺少的。

终止密码子 UGA 不但能以色氨酸的形式发生通读，而且能够作为硒半胱氨酸和磷酸丝氨酸的密码子而不是作为终止密码子。从大肠杆菌一直到人类的蛋白质中都发现了硒半胱氨酸(selenocysteine)。早先人们推测硒半胱氨酸是通过翻译后修饰而产生。现在已经知道，硒半胱氨酸是翻译时掺入到肽链中去的。大肠杆菌中 *selC* 基因编码一个特殊的 tRNA，有 95 个核苷酸，是最大的 tRNA 分子之一；其反密码子为 5′ UCA 3′，与 UGA 完全配对；接受臂上有 8 对配对的碱基，而不是通常的 7 对；在 D 嚼噗上有 5 个不同于一致序列的核苷酸，即 $G^9$, $G^{12}$, $U^{15}$, $C^{16}$, $C^{25}$。这个 tRNA 所荷载的是 L-丝氨酸，而不能荷载游离的硒半胱氨酸。荷载的 L-丝氨酸在 *selA* 和 *selB* 基因产物(分别为 50KDa 和 70 KDa) 催化之下变为硒半胱氨酸，然后在翻译时渗入肽链。UGA为什么在这种情况下不作为终止密码子而作为硒半胱氨酸的密码子被识别？其机制还不清楚。曾有不少人提出"上下文"(context)假说(即 $UGAR_n$)，但现在发现很多生物的硒半胱氨酸密码子 UGA 后并不一定是嘌呤。可能与生理条件有关，人们发现 *E. coli* 只有在无氧条件下才表达含有硒半胱氨酸的甲酸脱氢酶。

由上可见，在现代生命中UGA 是一个灵活多变的遗传密码。为了保持 UGA 密码子的灵活性，就必须严格控制细胞内释放因子 RF2 的含量(见第八章第七节)。如此之多的灵活的解读机制不可能是使用较原始遗传密码的早期生命所具备的。在这样的生命体系中，UGA 只可能简单地作为色氨酸的密码子之一。

再从线粒体的四个起始密码子来看，AUU 和 AUC 可能比 AUG 和 AUA 更为古老，因为它们更符合 RNY 假说(见下述)。根据以上各种事实，推测人类线粒体遗传密码比所谓通用密码更为古老，似乎是合理的。然而，这仍然不能排除流线化的可能性。

上面我们论述了线粒体遗传密码的古老性，但是线粒体密码并不是原始遗传密码，可能是原始遗传密码到现代通用密码之间的某个较晚阶段的遗传密码。20 种氨基酸到底什么时候和怎样与它们的密码子建立联系的问题，人们还只能进行推测。在密码形成之初，可能并没有 20 种氨基酸。人们通过对许多蛋白质基因的序列分析，发现其编码区中的嘌呤和嘧啶的相对位置有相当的恒定性，特别是密码子中 RNY 形式占优势。其中 R 为嘌呤核苷酸，Y 为嘧啶核苷酸，N 为四种核苷酸中的任意一种。这就暗示着，RNY 可能是最为古老的密码子。由古老的 RNY 密码子组成的 mRNA 中，可能只存在着这一种阅读框，因为当时还没有演化出翻译的起始信号和控制起始的因子。有趣的是，按照通用密码表(或线粒体密码表)，由 RNY 所编码的 8

种氨基酸（甘氨酸、丙氨酸、丝氨酸、苏氨酸、天门冬氨酸、天门冬酰胺、缬氨酸和异亮氨酸）很像生命前所存在的氨基酸，也经常在陨星中发现这些氨基酸，对酵母的3-磷酸甘油醛脱氢酶基因所进行的 RNY 周期分析表明，只有实际编码的阅读框与 RNY 周期符合程度最高，而另外两个阅读框的符合程度要低得多(图 7-6)。其他许多基因中，由于碱基突变(特别是插入或缺失)，使其 RNY 周期性不太明显，但在基因的某些区段仍然存在着这样的 RNY 周期。泛素是真核生物中高度保守的 76 个氨基酸的蛋白质。酵母的泛素基因 UBI1，UBI2，UBI3 的 76 个密码子符合 RNY 周期的均为 33 个，而 +1 阅读框和 +2 阅读框中符合 RNY 周期的则分别为 10 和 16 左右。如果在 *E. coli* 16S rRNA 和 23S rRNA 基因中寻找 RNY 周期，则会发现这两个基因中的嘌呤和嘧啶是随机分布的，没有 RNY 周期。这就表明，这两个基因从来就没有编码过蛋白质。

在密码的早期进化过程中，为什么首先采取 RNY 形式呢？首先 RNY 形式是一种自身互补的形式。其次，根据数百种 tRNA 分子的分析结果，发现 tRNA 分子上有重复的 RNY 形式。人们普遍认为，原始的 tRNA 具有双重功能，既接受氨基酸，又作为 mRNA。因此，在核糖体出现之前，只存在着 tRNA 这一类 RNA 分子，它们负责基因表达的全过程。当然，这个过程与现代生命中基因表达的全过程相比要简单得多，效率也要低得多。

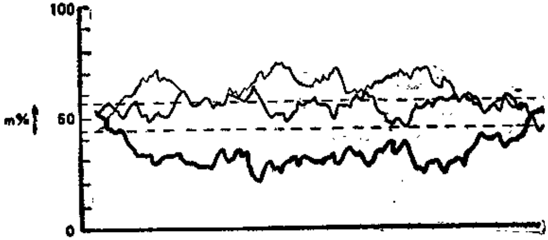



图 7-6　酵母 3-磷酸甘油醛脱氢酶基因中的三种不同阅读框的 RNY 信息。横

亚基，其中大亚基大约是小亚基的两倍大小。rRNA还直接参与核糖体与mRNA和tRNA的结合。核糖体中的蛋白质，rRNA以及其他的辅助因子在一起提供了翻译过程所需的全部酶活性。任何孤立的一个蛋白质或一组蛋白质或一组蛋白质与rRNA都不能产生任何酶活性。也就是说，这些酶活性只有在核糖体整体结构存在的情况下才会具备。因此核糖体蛋白质和rRNA都不是单独参与蛋白质合成反应的。

**表7-5 核糖体组成情况**

| | 大肠杆菌核糖体 | | | 大鼠肝细胞细胞质核糖体 | | |
|---|---|---|---|---|---|---|
| | 核糖体 | 小亚基 | 大亚基 | 核糖体 | 小亚基 | 大亚基 |
| 沉降系数 | 70 S | 30 S | 50 S | 80 S | 40 S | 60 S |
| 质量(道尔顿) | 2 520 000 | 930 000 | 1 590 000 | 4 420 000 | 1 400 000 | 2 820 000 |
| RNA 主要 | | 16 S=1 541碱基 | 23 S=2 904碱基 | | 18 S=1 900碱基 | 28S=4 700碱基 |
| 次要 | | | 5 S=120碱基 | | | 5.8 S=160碱基<br>5 S=120碱基 |
| RNA质量 | | 560 000 | 1 104 000 | | 760 000 | 1 820 000 |
| RNA比例 | 66% | 60% | 70% | 60% | 50% | 65% |
| 蛋白质 | | 21个 | 34个 | | 33个 | 49个 |
| 蛋白质质量 | | 370 000 | 487 000 | | 700 000 | 1 000 000 |
| 蛋白质比例 | 34% | 40% | 30% | 40% | 50% | 35% |

任何生物的核糖体都是由大小两个亚基组成。现将大肠杆菌核糖体和大鼠肝细胞核糖体的蛋白质组分和RNA组分列于表7-5。大肠杆菌核糖体共有55个蛋白质，34个在大亚基中，以L表示；21个在小亚基中，以S表示。23S rRNA和5S rRNA在大亚基中，16 S rRNA在小亚基中。其中L26与S20完全相同，实际上L26并不是每个大亚基都有的，大约每5个大亚基才有一个L26。L7是L12的氨基末端乙酰化的形式，实质上是同一种蛋白质。L8是由L10，L7，L12各一个分子组成的复合物。除此之外，其他蛋白质均各不相同。真核生物有82个核糖体蛋白质，49个在大亚基中，33个在小亚基中，亦分别用L和S表示。核糖体RNA 28S，5.8S和5S在大亚基中，而18S rRNA在小亚基中。真核生物L7/L12蛋白质与原核生物的L7/L12蛋白质的序列高度相关，在体外组建核糖体时可以互换。在核糖体蛋白质中，大多数是碱性蛋白质，有的蛋白质中Arg+Lys占氨基酸总数的三分之一。这是这些蛋白质能与RNA结合的重要原因。

尽管原核生物和真核生物的核糖体蛋白质和rRNA差异很大，但核糖体的总体结构却很相似，特别是负责与mRNA结合的小亚基更是如此。原核生物和真核生物的rRNA都有甲基化现象。大肠杆菌16S rRNA含有10个甲基化核苷酸，23S rRNA含有20个甲基化核苷酸。这些甲基化大多发生在A和G上。哺乳动物18S和28S rRNA分别含有43个和74个甲基化位点，几乎全部发生在核糖的2′~0上。已知这些甲基化与核糖体RNA的转录处理过

程中酶的识别有关(见第六章)，与核糖体功能的关系还不清楚。不同生物的 rRNA 序列也有一定的相关性，特别是甲基化位点和形成二级结构的配对碱基。原核生物的 5S rRNA 和真核生物的 5.8 S rRNA 分子都有许多配对的碱基，因此其序列亦高度保守。因此，在研究生物进化上，常用 5S(5.8S)rRNA 的碱基序列差异来推算进化的年代。

## 二、核糖体的装配

各种核糖体蛋白质在大小亚基内都有一定的位置，rRNAs 是其中的骨架。有的核糖体蛋白质直接与 rRNA 结合，在这些蛋白质与 rRNA 结合以后改变了 rRNA 的构象，从而使另外一些蛋白质再与 rRNA 结合。有些蛋白质不直接与 rRNA 接触，而是结合在其他蛋白质上。

人们对大肠杆菌核糖体 30S 亚基的装配过程研究得比较清楚。首先将 16S rRNA 与一组大约 15 种小亚基蛋白质 (S4、S8、S20、S7、S15、S13、S16、S17、S9、S19、S6、S18、S11、S5 等)在低温下(0℃)相互作用，则形成一种 R1 颗粒，然后将温度提高至 40℃，R1 颗粒就改变了构象，变为 21S 的 R1* 颗粒，然后加入其余 6 种蛋白质，则生成 30S 亚基。整个装配过程是自发的，不需要酶的催化，只是需要提高温度。如果将 EDTA 加到 30S 亚基体系中，由于失去了 $Mg^{++}$，30S 亚基就释放出若干种蛋白质(split protein)而变为一种类似于 R1 的颗粒(图7-7)

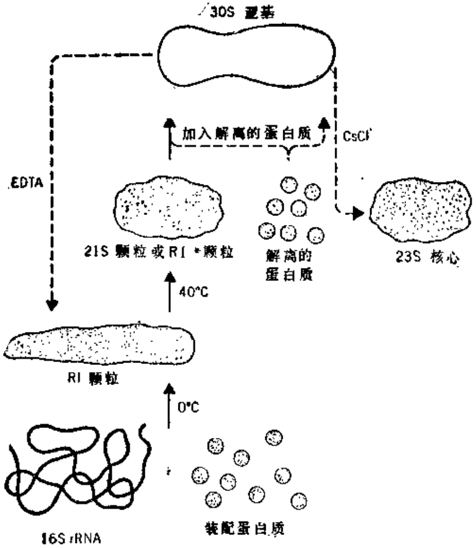



图 7-7 30S 亚基的装配步骤

上述体外重建核糖体的试验需要提高温度，就暗示着能找到在低温下不能装配核糖体的冷敏感突变体。根据这一设想，Nomura等人分离到若干种*sad* (subunit assembly defective)突变体。这些突变产物，有的是核糖体蛋白质，有的是直接或间接影响亚基装配的成分。*sad*

突变体细胞中，积累了较小的核糖体亚基的前体，其中包含 rRNA 和部分核糖体蛋白质。根据 *sad* 突变体的研究，已发现小亚基有 26S 和 21S 前体，大亚基有 43S 和 32S 前体。可能还会有其他前体尚未被发现。这些前体的存在表明核糖体亚基的装配存在着若干重要的步骤。

## 三、核糖体突变

突变体的研究，不仅是了解亚基装配步骤的重要手段，而且也是了解核糖体内各亚基功能和相互作用的重要手段。其原理我们已在第五章中讲过。同样，这些突变体只可能是温和的突变体和条件体致死突变体。

选择和鉴定细菌的核糖体突变有许多途径。一种是选择对抑制核糖体功能的抑制剂不敏感的突变体。例如链霉素能与小亚基的 S12 蛋白结合从而抑制蛋白质的合成。而突变体的 S12 蛋白失去了与链霉素结合的特性，从而表现出对链霉素的抗性。目前已通过这类方法选择出5种小亚基蛋白质和 4 种大亚基蛋白质的突变体。利用这些突变体可以研究有关的核糖体蛋白质的功能。

另一种选择途径是直接选择影响蛋白质合成精确度的突变体。这一突变效应有时又可以为另一个核糖体蛋白质的突变所抑制(基因间间接抑制突变)。这样，就可以研究亚基组分之间的相互作用。目前用这种方法已经获得了 6 种核糖体蛋白质的突变体。

上面已经讲过 *sad* 突变体直接影响到核糖体亚基的装配，表现出对低温的敏感。目前已分离到的 *sad* 突变体涉及到 4 种小亚基蛋白质和 1 种大亚基蛋白质。核糖体蛋白质也同样存在着温度敏感突变，使细胞在高温下生长缓慢或停止生长，这是因为核糖体蛋白质在高温下失去原有的功能。

春日霉素(kasugamycin)的抗性突变体表明了 16S rRNA 在蛋白质合成起始过程中的作用。春日霉素能抑制蛋白合成的起始，这是因为春日霉素能与 16S rRNA 靠近 3′ 末端的保守序列 G-m2 $^6$A—m2 $^6$A 相结合，而突变体 *ksgA* 的甲基化酶基因突变，不能使这相邻的两个 A 发生甲基化。当春日霉素结合于这两个甲基化的腺苷酸时，fMet-tRNA$_f$ 就不能与之结合；而突变体的 16S rRNA 的这一部分保守序列，既不能与春日霉素结合，在50S亚基不存在时也不能与 fMet-tRNA$_f$ 结合。但是当突变体的 30 S 亚基和 50 S 亚基同时存在时能低效率地结合 fMet-tRNA$_f$，仍能使蛋白质起始合成。这就表明，16S rRNA 的 3′ 端序列是与 50S 亚基以及 fMet-tRNA$_f$ 的结合有关。50S 亚基的某些蛋白质的突变体使得核糖体必须依赖于春日霉素。这也是基因间间接抑制突变的例子。某些细菌所产生的大肠杆菌素 E$_3$ 能够切下 16S rRNA的 3′ 末端49个碱基的序列，则完全抑制了蛋白质合成的起始，这也说明了 16 S rRNA 的 3′末端序列在翻译起始过程中的作用。

## 四、核糖体的活性位点

在核糖体上，我们能找到若干为蛋白质合成所不可缺少的活性位点。起始翻译时，30S亚基与 mRNA 以及与起始 tRNA 和起始因子构成的复合物相结合，然后再与 50 S 亚基相结合而成为 70S 核糖体。70S核糖体上有 A 位和 p 位，A 位供氨酰基-tRNA 结合，p 位供肽基-tRNA 结合。肽基转移酶活性位点就在 50 S 亚基上，位于 p 位和 A 位之间。此外，还有转位因子 EF-G，延伸因子 EF-Tu 和 5 S rRNA 位点等，这几个位点均在 50 S 亚基上。人们采用亲和标记、交联剂、抑制剂、抗体等来分析这些活性位点的结构和功能。下面是主要的活性位点以及

所涉及的核糖体蛋白质(图 7-8)。

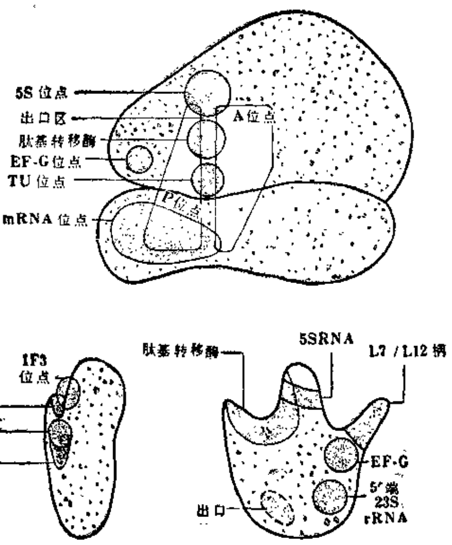



图 7-8 核糖体的主要活性位点

1. mRNA 结合位点 位于 30 S 亚基的头部。30 S 亚基与 mRNA 的起始结合，必须有有功能的 S1 蛋白。S1 蛋白具有狭长的形状，与单链核酸有高度的亲和力。S1 蛋白与 mRNA 的结合，防止 mRNA 链内碱基对的形成。这种链内碱基对不利于翻译作用的进行。通过亲和标记实验，发现与 S1 蛋白直接有关的蛋白质有 S18 和 S21。这三种蛋白质构成一个小的结构域，负责与 mRNA 的初始结合以及与起始 tRNA 的结合。这一位点也是 IF3 的结合位点。这一位点邻近 p 位点，这一位点可能还包括蛋白质 S 3、S 4、S 5、S 12、以及 16 S rRNA 的 3′ 末端区域。16S rRNA 3′ 端的富含嘧啶区与 mRNA 上 AUG 之前的富含嘌呤区发生碱基配对，是 30S 亚基与 mRNA 的初始结合不可缺少的。

2. P位点 通过亲和标记实验，发现 P 位点大部分位于 30S 亚基，小部分位于 50S 亚基。16 SrRNA 的 3′ 末端区域也是 P 位点不可缺少的组成部分。大亚基上涉及的重要蛋白质有 L2 L27 以及 L14、L18、L24、L33。P 位点能够与起始 tRNA(fMet-$tRNA_f$)相结合。

3. A 位点 A 位点靠近 P 位点，但其精确位置尚不清楚。16 S rRNA 是其构成成分之一。A 位点主要在 50S 亚基上，涉及的核糖体蛋白质有 L1、L5、L7/L12、L20、L30、L33。

4. 肽基转移酶活性位点 位于 P 位和 A 位的连接处，靠近 tRNA 的接受臂。涉及的核糖体蛋白质有 L2、L16 以及 L3、L4、L15，还涉及到 23 S rRNA。

5. 5S rRNA 位点 在 50S 亚基上，靠近肽基转移酶位点。涉及的蛋白质有 L5、L8、L25。tRNA 的 TψC 噜噗的保守序列与 5S rRNA 的序列互补，可能与 tRNA 的进入有关，但还缺乏

有力的证据。

6. EF-Tu 位点　位于大亚基内，靠近 30S 亚基，涉及的蛋白质有L5、L1、L20和L7/L12。这一位点与氨酰基-tRNA 的结合有关。

7. 转位因子 EF-G 结合位点　在大亚基上，靠近与小亚基的界面处。这一位点靠近 S12，并与 L7/L12 二聚体相邻。

以上位点的空间关系还有许多问题有待解决。例如 P位点和 A位点必须很靠近，因为两个 tRNA 要与相邻的两个密码子配对。相邻的两个密码子的最大距离为 10Å，而每个 tRNA的直径就有 20 Å。可能是 mRNA 在两个密码子之间发生扭转，使两个 tRNA 从不同的方向与 mRNA 结合。如果是这样，tRNA 从 A 位轻轻地弹出，转过一个角度进入 P位。因此 mRNA 对于核糖体的移动实质上是 tRNA 的移动。因为反密码子和密码子之间的配对，才能使每次移动的距离是精确的一个密码子。

## 第三节　肽链的合成[24, 28~36]

细胞内的数十种 tRNA 分子构成了氨基酸的运输大军，而运来的氨基酸在核糖体内迅速地装配成多肽链。在这个过程中所需的能量由 GTP 水解提供。这个装配过程可以分为三个阶段：(1) 起始阶段，第一个肽键形成之前的各步反应。它需要核糖体结合于 mRNA，从而促使包括第一个氨酰基-tRNA 在内的起始复合物的形成，这是一个相对缓慢的步骤。(2) 延伸阶段：包括从第一个肽键形成到最后一个肽键形成之间的全部反应。这是一个相当迅速的过程。(3) 终止：包括已完成的肽链和核糖体的解离。这一过程要比延伸过程中加上一个氨基酸所需的时间长得多。

蛋白质合成过程总的说来仍然是一个非常迅速的过程。在 37℃ 时，细菌细胞内合成速度为每秒 15 个氨基酸。所以合成一个 300 个氨基酸的蛋白（相当于蛋白质的平均分子量）只需要 20 多秒钟。细菌细胞内，80% 的核糖体都在忙于蛋白质的合成，很少有游离的核糖体。真核生物中蛋白质合成速度要低一些。红血球蛋白质合成速度只有每秒 2 个氨基酸。

### 一、合成的起始

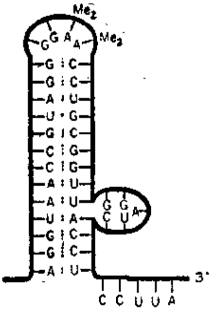
1. 起始 tRNA 与起始密码子的识别　合成的起始密码子为 AUG（在细菌中偶尔也有用 GUG 的）。在细菌中，约有 50% 的蛋白质的 N 端为 Met。后来人们发现了 N-甲酰-甲硫氨酰-tRNA，其氨酰基的氨基末端已被甲酰化，使之只能用于起始而不能用于延伸过程。

细菌中有两种 tRNA 能够携带甲硫氨酸。一种是 $tRNA_f^{Met}$，它只能识别起始密码子 AUG，一种是$tRNA_m^{Met}$，它只能识别内部 AUG 密码子。$tRNA_f^{Met}$ 首先与 Met 结合，然后在 Met 的 $NH_2$ 上产生甲酰化作用从而封闭了这个氨基。这个氨酰基-tRNA 可缩写为fMet-$tRNA_f$。起始密码子 AUG 和 GUG 均由这种 $tRNA_f$ 所识别。而这两个密码子如果在 mRNA 的内部则分别由 $tRNA_m^{Met}$ 和 $tRNA^{Val}$ 所识别。而在真核生物中，情况有所不同。起始密码子只能是AUG，而没有GUG作为起始密码子的。识别起始AUG与内部AUG的工作同样是由不同的 tRNA 担任的，负责起始 AUG 识别的是 $tRNA_i^{Met}$，负责内部 AUG 识别的是 $tRNA_m^{Met}$。但 Met-$tRNA_i$ 并不被甲酰化。在哺乳动物线粒体中同样有两类 tRNA 分别识别起始 AUG 和内部 AUG，而且起始 tRNA 上携带的是甲酰化的甲硫氨酸。不同的是线粒体 Met-$tRNA_f$

表 7-6 *E.coli* 的蛋白质起始合成需要三种起始因子

| 因子 | 质量(KDa) | 分子数/自由核糖体 | 功能 |
|---|---|---|---|
| $IF_3$ | 23 | 25% | 结合 mRNA 与亚基解离 |
| $IF_2$ | 100 | ? | 起始 tRNA 的结合 |
| $IF_1$ | 9 | 15% | 可能为 IF2 的循环因子，促进 IF2 的解离 |

(a)



16S rRNA

(b)

| | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| 1 | CUG | AGU | AUA | AGA | GGA | CAU | AUG | CCU | AAA | UUA |
| 2 | CUU | UGG | GUC | AAU | UUG | AUC | AUG | GCA | AAA | UUA |
| 3 | UUA | CUA | AGG | AUG | AAA | UGC | AUG | UCU | AAG | ACA |
| 4 | AUG | UAC | UAA | GGA | GGU | UGU | AUG | GAA | CAA | CGC |
| 5 | UUU | AAU | CGA | AAC | UUC | CUC | AUG | AAA | AAG | UCU |
| 6 | AAU | CUU | GGA | GGC | UUU | UUU | AUG | GUU | CGU | UCU |
| 7 | UUG | CUC | GAG | GCC | UCC | ACU | AUG | AAA | UCG | CGU |
| 8 | AGG | UCU | AGG | AGC | UAA | AGA | AUG | GAA | CAA | CUC |
| 9 | GCG | UUG | AGG | CUU | GCG | UUU | AUG | GUA | CGC | UGG |
| 10 | AUC | UAG | AGG | GUA | UUA | AUA | AUG | AAA | GCU | ACU |
| 11 | GGC | AUG | ACA | GGA | GUA | AAA | AUG | GCU | AUC | G |
| 12 | AGC | CUA | AUG | GAG | CGA | AUU | AUG | AGA | GUU | CUG |
| 13 | CCC | GAU | UAA | GGA | ACG | ACC | AUG | ACG | CAA | UUU |
| 14 | CAA | UUC | AGG | GUG | GUG | AAU | GUG | AAA | CCA | GUA |
| 15 | UUC | ACA | CAG | GAA | ACA | GCU | AUG | ACC | AUG | AUU |
| 16 | CAU | CAA | GGA | GCA | AAG | CUA | AUG | GCU | UUA | AAU |
| 17 | UAU | UCA | GGA | ACA | AUU | UAA | AUG | UCU | AUC | ACU |
| 18 | AGC | GAG | CUG | AGG | AAC | CCU | AUG | GUU | UAC | UCC |

16S 3' end $_{HO}$AUUCCUCCACUAG-5'

图 7-9

(a) 原核生物 16S rRNA 3' 末端的碱基序列及其发夹结构 (b) 原核生物基因的S-D序列：
1. 噬菌体 Qβ A 蛋白； 2. Qβ 衣壳蛋白； 3. Qβ 复制酶； 4. 噬菌体 λ Cro 蛋白；5. 噬菌体 f1 衣壳蛋白；6. 噬菌体 φX174A； 7. φX174A*； 8. φX174B； 9. φX174E； 10. *E.coli* 脂蛋白； 11. RecA； 12. GalE； 13. GalT； 14. LacI； 15. LacZ； 16. L10； 17. L7/L12； 18. RNA聚合酶β亚基

还能将 AUA，AUC，AUU 用作起始密码子。

N 末端的甲酰基，一般在合成 15～30 个氨基酸后由去甲酰基酶除去。

2、起始复合物与 70S 核糖体的形成　形成起始复合物，除了需要 mRNA，荷载了甲硫氨酸的起始 tRNA，30 S 亚基、50 S 亚基之外，还需要起始因子。在细菌中，有三种起始因子（表 7-6）。这些因子并不牢固地结合于核糖体，在起始复合物形成之后就很快解离。

起始复合物与 70S 核糖体形成的步骤如下：

(1) IF3 促进 70S 亚基的解离，并与 30S 亚基结合。结合有 IF3 的 30S 亚基与 mRNA 结合。在 mRNA 上起始密码子 AUG 上游方向 4～13 核苷酸之前有一段富含嘌呤的序列，其一致序列为 AGGAGG，(称为 Shine-Dalgarno 序列)，能够与 16 S rRNA 3′ 末端的富含嘧啶的序列(有时也不恰当地称为 16S rRNA 的 Shine-Dalgarno 序列)相结合(图 7-9)。不同基因的 Shine-Dalgarno 序列不完全相同，从而控制单位时间内起始复合物形成的数目，最终控制翻译产物的数量。30S 亚基在 mRNA 上的结合位置，正好使 30S 亚基上的部分 P 位对准起始密码子 AUG 以便携带甲酰甲硫氨酸的起始 tRNA 进入 P 位。

(2) 起始因子 IF2 结合于起始 tRNA，然后 IF2 再与 30S 亚基结合(或许 IF2 先与 30 S 亚基结合，再识别起始 tRNA)，使起始 tRNA 进入 30 S 亚基的部分 P 位。不管按什么顺序，IF2 都是这一阶段中 30 S 亚基不可缺少的部分。

(3) 在 IF2 与起始 tRNA 形成的二元复合物结合于 30S 亚基后，GTP 分子立即与 30 S 亚基结合。IF2 可能具有 GTP 酶的活性或者活化其他某种核糖体蛋白质使之具有 GTP 酶的活性。这时起始复合物已经完全形成。

(4) 50S 亚基结合上来，GTP 水解，释放出的能量改变了 30S 亚基和 50S 亚基的构象，促进 70S 核糖体的形成。同时释放出起始因子 IF3 和 IF2。IF1 可能促进 IF2 的释放。这时核糖体上的 P 位和 A 位都已处于正确的姿态，P 位已被起始 tRNA 充满，而 A 位也正在准备接受一个能与第二密码子配对的 tRNA。

起始复合物和 70S 核糖体形成的主要步骤示于图 7-10。

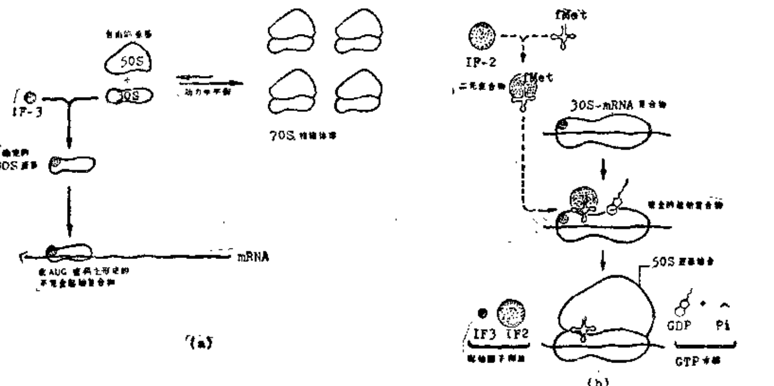



图 7-10　起始复合物和70S核糖体的形成过程

(a) IF-3 促进 70S 核糖体的解离与不完全起始复合物的形成

(b) IF-2 促进完全起始复合物和 70S 核糖体的形成

3．真核生物蛋白质合成的起始　真核生物的起始复合物并不是在起始密码子AUG处形成，而是首先在mRNA的5′末端形成，其识别信号就是5′末端的帽子结构。在poly(U)的5′末端加上帽子结构，也能够形成起始复合物，说明帽子结构这一信号足以使起始复合物得以形成。在帽子处形成的起始复合物，然后沿着mRNA移动，直到发现起始密码子AUG，60S亚基结合上来，才形成80S核糖体。其过程与原核生物有许多相似之处。然而，真核生物的起始因子更多，目前已发现有9种起始因子(表7-7)，其中符号eIF表示真核生物起始因子。这些因子的作用细节大多不太清楚。

表7-7　真核生物蛋白质合成起始因子

| 因子 | 结构 | 功能 |
|---|---|---|
| eIF3 | >500Kd，多聚体 | 结合mRNA |
| eIF1 | 15Kd，单体 | 协助mRNA的结合 |
| eIF4B | 单体 | 协助mRNA的结合 |
| eIF4A | 单体 | 协助mRNA的结合以及结合ATP |
| eIF6 | 23Kd，单体 | 防止40S亚基与60S亚基结合 |
| eIF5 | 150Kd，单体 | 释放eIF2和eIF3结合60S亚基 |
| eIF4C | 单体 | |
| eIF4D | 单体 | 不清楚 |
| eIF2 | $\alpha$，35Kd<br>$\beta$，38Kd　三聚体<br>$\gamma$，55Kd | 结合GTP，受磷酸化作用控制<br>促进eIF2循环的因子结合<br>Met-tRNA |

真核生物起始复合物的形成过程是：GTP首先与eIF2结合，这一结合就增加了eIF2与起始tRNA的亲和力，然后三者便结合成一个三元复合物。三元复合物直接与40S亚基结合，这一过程不需要mRNA的存在。40S亚基-三元复合物在eIF3的存在下与mRNA结合。在此过程中水解1分子ATP以提供能量。这样，起始复合物就形成了(图7-11)。然后起始复合物沿着mRNA向起始密码子AUG移动以便形成80S核糖体。

## 二、延　伸

1．转肽与肽键的形成　当完整的核糖体在起始密码子上形成以后，P位已被携带(甲酰)甲硫氨酸的起始tRNA所充满，而A位仍然空着。如果不考虑密码子和反密码子配对的话，则除起始tRNA之外的任何一种tRNA都可以进入A位。然而，tRNA进入A位不是自发进行的，而是需要延伸因子EF。延伸因子有三种，一种是热不稳定的，叫做EF-Tu；一种是热稳定的，叫做EF-Ts；第三种是依赖于GTP的延伸因子EF-G，又称转位因子。EF-Tu首先与GTP结合，然后与氨酰基-tRNA结合形成三元复合物。这样的三元复合物才能够进入核糖体的A位。其中GTP的存在是氨酰基-tRNA能够进入A位的先决条件之一，但并不需

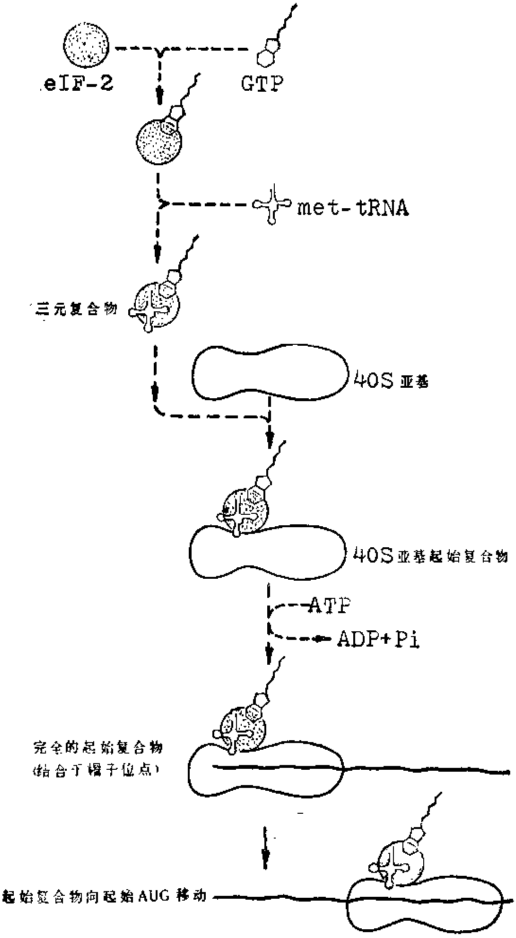



图 7-11 真核生物蛋白质合成起始复合物的形成

**GTP 的水解**。一旦进入 A 位以后，GTP 立即水解成 GDP，EF-Tu、GDP 这个二元复合物就与氨基酰-tRNA 解离而释放出来。只有这时，肽基转移酶才能把位于 P 位的甲酰甲硫氨酰基或肽基转移到 A 位的氨酰基-tRNA 的氨基上并形成第一个肽键或新的肽键。这一步的关键是 GTP 的水解和 EF-Tu·GDP 的释放。

现在已有充足的证据说明 GTP 的存在及其水解释放与氨酰基-tRNA 进入 A 位点和肽键形成的关系。人们使用一个不能被水解的 GTP 同系物 GMP-PCP，其中 C 代表连接 $\beta$ 和 $\gamma$ 磷原子的亚甲基而不是 GTP 中的氧原子。利用这种同系物照样能够形成三元复合物，然后

进入 A 位；但是肽键不能形成。这就说明三元复合物的形成及其进入 A 位点所需的是 GTP的存在而不是它的水解。GTP 水解后，EF-Tu·GDP 就不能有效地与氨酰基-tRNA 结合，这说明鸟嘌呤核苷酸控制着 EF-Tu 的分子构象。对于肽键的形成，不但需要 GTP 的水解，而且需要 EF-Tu·GDP 的释放。人们用黄霉素 (kirromycin)来抑制 EF-Tu 的功能。当 EF-Tu 与黄霉素结合以后，照样能够形成三元复合物并使氨酰基-tRNA 进入 A 位，GTP 也照常水解，但 EF-Tu·GDP 不能从核糖体上释放下来。其继续存在妨碍了肽基-tRNA和氨酰基-tRNA之间肽键的生成。其结果是蛋白质合成停滞不前。

2. *转位*　当肽键在 A 位上形成之后，也只有在这个时候，转位因子 EF-G 和 GTP 才能够结合上来。EF-G 与 GTP 可能形成很松弛的复合物，也可能二者在肽键形成之后直接结合到核糖体上来。同样，这种结合需要的是 GTP 的存在而不是 GTP 的水解。因为 GMP-PCP 照样能促进这种结合。然后，由核糖体中具有 GTP 酶活性的某种蛋白质将 GTP 水解，在 A 位生成的肽基-tRNA 才能转移到 P 位同同时将原来P位上空载的 tRNA 逐出核糖体，mRNA 也移动了一个密码子。在转位以后，EF-G 和 GDP 必须释放出来，下一个氨酰基-tRNA 的三元复合物才能进入 A 位。有一种抗生素叫甾酸霉素 (fusidic acid) 能够使核糖体停在转位后的状态。这是因为甾酸霉素稳定了 EF-G·GDP 与核糖体的复合物，使下一个氨基酸不能加到肽链上来。

3. *两类延伸因子的交替作用*　由上可见，只有 EF-Tu 离开核糖体以后，EF-G 和 GTP 才能结合上来；同样，只有 EF-G 离开核糖体后，新的氨酰基-tRNA 三元复合物才能进入 A 位。现在还不知道延伸因子互相排斥的现象是否由于这些因子对核糖体上互相重叠的结合位点的竞争而引起的；人们已知一种药物硫链丝菌素 (thiostrepton) 能抑制两种延长因子的结合。然而有一点是必须的，那就是 P 位点和 A 位点的状态。只有当 P 位点被肽基 tRNA 占据而 A 位点空着时，氨酰基-tRNA 三元复合物才能进入 A 位点；只有当肽键生成后，肽基转移到 A 位的氨酰基-tRNA 上，P位tRNA空载，EF-G 和 GTP 才能结合到核糖体上来。加上鸟嘌呤核苷酸控制着这两种延伸因子构象，从而改变了它们对于核糖体的亲和力，使一种因子与核糖体的结合与释放有序地进行。这样，两种延伸因子就能交替地与核糖体发生相互作用，使肽键合成和转位作用有条不紊地进行，整个延伸过程表示于图 7-12。

4. *Ts 循环*　延伸因子 EF-Ts 的作用就在于使使用过的 EF-Tu·GDP 能够转变为有用的形式 EF-Tu·GTP。首先 EF-Ts 将 GDP 置换掉，生成结合的因子 EF-Tu·EF-Ts。这就是人们最初发现的 T 因子即转移因子。然后 EF-Ts 又被GTP置换出来而生成EF-Tu·GTP，而 EF-Ts 又可进行下一次循环。Ts 循环过程示于图 7-12(a)。

### 三、终止和肽链的释放

无论原核生物还是真核生物都有三种终止密码子 UAG (amher)、UAA(ochre)和 UGA (opal)。没有一个 tRNA 能够与终止密码子作用，而是由特殊的蛋白质因子促成终止作用。这类蛋白质因子叫做释放因子。原核生物有三种释放因子：RF1 识别 UAA 和 UAG；RF2 识别 UGA 和 UAA；RF3 能够刺激 RF1 和 RF2 的活性。在真核生物中，只有一种释放因子 eRF。eRF 需要 GTP 与之结合才能结合于核糖体，GTP 可能在终止反应后被水解，这一作用可能与 eRF 与核糖体的解离有关。

不管原核生物还是真核生物，释放因子都是作用于 A 位点，并且需要 P 位被肽基-tRNA

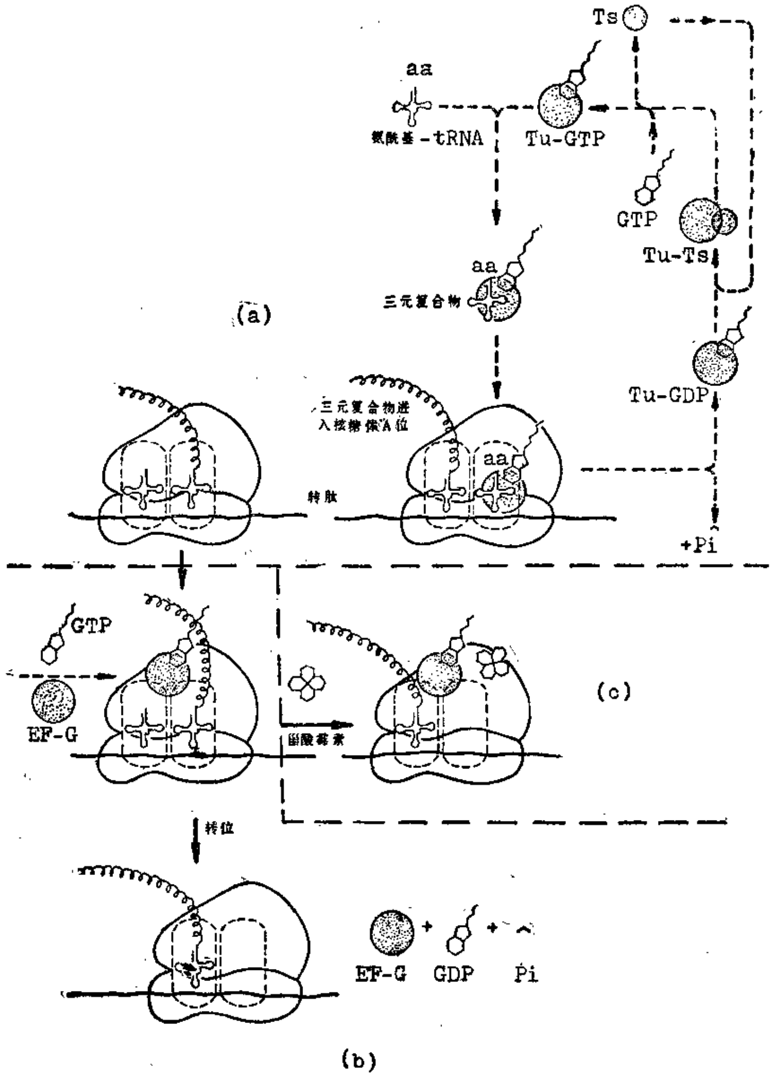



图 7-12　肽链的延伸过程

(a) 肽链的形成和 Ts 循环；(b) 转位；(c) 甾酸霉素将核糖体停在转位后状态

所占据。这个过程同样需要肽基的转移以及把空载的 RNA 逐出核糖体。肽基的转移仍然是由肽基转移酶活性起作用的，只不过由于释放因子的某种作用使肽基转移到水分子上因而并不形成新的肽键。核糖体是如何从 mRNA 上释放下来的，现在还不清楚，可能是释放因子触发了核糖体构象的改变，并有可能涉及到未知的蛋白质因子。

## 第四节 mRNA 的结构与翻译[28,31,37~40]

上面我们已经讲到了原核生物 mRNA 上的 Shine-DaLgarno 序列，真核生物的帽子结构以及两种 mRNA 的起始密码子 AUG 都是蛋白质起始合成所不可缺少的。而 mRNA 上的 UAG，UAA 和 UGA 三种终止密码子又是肽链的终止和释放所不可缺少的。除此之外，mRNA 上还有许多结构特点与翻译过程密切相关。

### 一、原核生物的多顺反子 mRNA 的翻译

原核生物的 mRNA 大部分为多顺反子。在各个顺反子即基因之间，常有一段非编码的间隔区(spacer)。不同的间隔区长度变化很大，可以从 100 个碱基左右直到 1 个碱基，甚至有前一个基因的终止密码子与后一个基因的起始密码子相互重叠甚至编码序列相互重叠的现象(如我们在第二章中讲到过的 $\phi$X174 的基因组织情况)。如果基因间的间隔序列较长，则核糖体在第一个基因的终止密码子发生解离，并脱离了 mRNA。因此第二个基因产物合成的起始是独立的，同样需要全部的起始过程。如果基因间间隔序列的长度只相当于正常的 S-D 序列到起始密码子的距离并具有 S-D 序列特征，则核糖体大小亚基分离后，小亚基并不离开 mRNA，大亚基暂时离开，但又会很快地结合上来。如果前一基因的终止密码子与后一基因的起始密码子部分重叠，则核糖体两个亚基可能都不离开 mRNA 而接下去翻译第二个基因产物。因而在原核生物的多顺反子 mRNA 中，各基因产物的数量并不一定都相等，特别是基因间间隔序列较长的情况下更是如此。这时候，各基因 S-D 序列则起关键作用。乳糖操纵元中各基因产物数量就不相等。半乳糖苷酶(Z 基因产物):半乳糖苷透性酶(Y 基因产物):半乳糖苷乙酰化酶(A 基因产物)大约为 1:0.5:0.2。造成这种结果可能还有其他的原因，如后面基因迟转录出来，而且 mRNA 的 3′ 末端更容易受到核酸酶的进攻而降解等。如果上游基因发生无义突变，则可以在转录水平上引起突变的极性效应(后面基因无转录产物)，这在第六章中已经讲过。这种无义突变并不都造成转录的过早终止，突变的极性效应也可能发生在翻译水平上。因为无义突变点下游序列没有被核糖体占据，就有可能与下一个基因 mRNA 的核糖体结合位点形成二级结构，从而使下游基因不能进行翻译或降低了翻译效率。

### 二、真核生物核糖体形成位点与双功能 mRNA

真核生物的 18S rRNA 与原核生物的 16S rRNA 在其3′末端有很大的同源性(图 7-13)。

由图 7-13 可见，由于 18S rRNA 缺少 CCUCC 序列(与 mRNA 上的 S-D 序列互补)，因此真核生物 mRNA 可能并不存在 S-D 序列。根据 Kozak 等人的研究，大多数起始密码子 AUG的合适的"上下文"为CCACC<u>AUG</u>G。在帽子结构上形成的 40S 亚基起始复合物向 3′方向移动时，如发现 AUG 密码子处于这样的"上下文"中，就不再向前移动，而是与 60S 亚基结合成为完整的核糖体，于是从 AUG 上开始合成蛋白质。如果这个 AUG 的上下文不大合适，

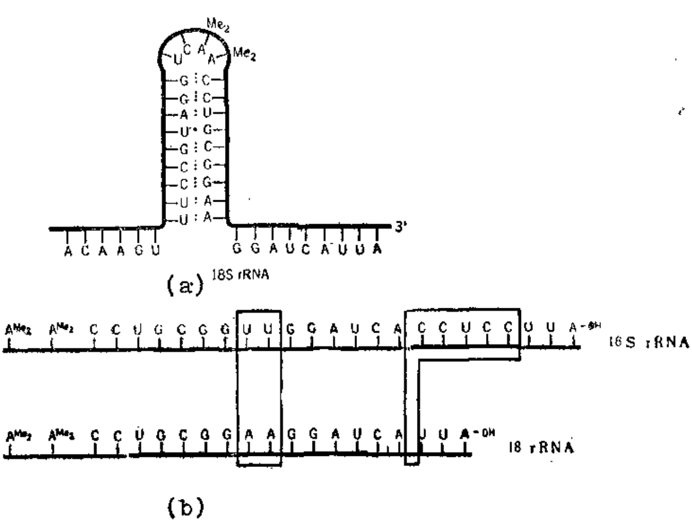



图 7-13 真核生物(哺乳动物)18S rRNA3′末端结构

(a) 3′末端发夹结构; (b) 原核生物 16S rRNA 与真核生物 18S rRNA 的序列比较。虚线框内示出二者碱基差异

有时 40S 亚基起始复合物就会继续向前移动，在碰到下一个位于较合适的上下文中的 AUG时停下来而形成80S亚基核糖体，并从这个AUG上开始合成蛋白质。这样，就有可能从两个 AUG 上合成两个蛋白质。如果两个 AUG 属于同一阅读框，则形成两个长短不同的蛋白质；如果两个 AUG 处于不同的阅读框中，则合成两个序列完全不同的蛋白质。这样，一条 mRNA 可以合成两种蛋白质，因此称为双功能 mRNA(bifunctional mRNA)。现已发现许多病毒都具有双功能mRNA，并已分离到这样的蛋白质。SV40 的衣壳蛋白VP2和VP3就是一个典型的例子。然而，在绝大多数的真核生物基因的表达过程中，总是识别第一个 AUG，这可能是由于真核生物 $tRNA_i^{Met}$ 反密码子 3′端的一个相邻碱基A 的烷基化阻碍了它与AUG 配对的摇摆性。

### 三、只有最后一个终止密码子的多基因 mRNA 的翻译

真核生物中广泛存在着一种序列高度保守的 76 氨基酸所组成的小分子蛋白质，这种蛋白质称为泛素(ubiquitin)。从昆虫直到人类，泛素的氨基酸序列全部相同。这类基因大多呈“尾-首”相连的形式组成多基因簇，中间无任何间隔序列(图 7-14)。酵母有 5 个这样的重复基因组成基因簇，人类则有9个。除第一基因前有启动子外，后面基因均无启动子，基因内启动子(不是该基因使用而是后一基因使用)的可能性也不存在。也就是说不可能每个基因单独转录，也无转录后处理所具备的特征。因此只能在翻译后将各基因产物分离。在细胞内实际上并不存在这样的多蛋白质 (5 聚体或 9 聚体)。因为细胞内存在着能够切除多蛋白质之间肽键的特殊的蛋白酶——泛素羧端水解酶。这种酶识别的是泛素的空间构象。当翻译进行到一个单体出来后不久(可能只比单体多几个氨基酸)，泛素分子的空间构象就能形成，这种酶就能识别而将单体切除下来。像这样的多蛋白质才是真正的多基因蛋白质，不过已发现的这类例子很少。

图 7-14 酵母泛素基因 UBI4 的结构：5 个基因(均无内元)尾-首相连，最后一个基因有一个附加的氨基酸(Asn)密码子

## 第五节 蛋白质在细胞内的越膜运输、定位和翻译后处理[41~53]

### 一、概 述

细菌不但合成在细胞内起作用的蛋白质，而且合成在细胞膜上起作用的蛋白质。事实上，革兰氏阴性细菌如大肠杆菌的细胞膜具有两层膜结构：内膜含有参与能量代谢和营养物质转运的蛋白质；外膜含有促进离子和营养物质进入细胞的蛋白质（这些蛋白质中有些被噬菌体用作受体）；在内膜和外膜之间的间隙称为周质(periplasm)，其中含有水解酶(如核酸酶，蛋白酶)以及营养物质结合蛋白。

真核生物的细胞结构更为复杂，不但细胞体积大，而且细胞内具有各种不同的细胞器，这些细胞器种类具有各自独特的膜结构。在细菌细胞内起作用的蛋白质一般是靠扩散作用而分布到它们的目的地。而真核细胞中蛋白质的定位要复杂得多，不但要决定是否越膜，而且要决定越过哪一种类的膜。

在越膜过程中，有时伴随着与翻译同时发生的处理过程 (cotranslational processing)。有些蛋白质在翻译完毕之后还要经受许多共价修饰，这个作用称为翻译后处理 (posttranslational processing)。

### 二、细菌中蛋白质的越膜

细菌的内膜蛋白、外膜蛋白和周质蛋白是怎样越过内膜而到达其目的地呢？绝大多数越膜蛋白都具有大约 15～30 个氨基酸的 N 端信号序列或信号肽。在信号序列的后半段以疏水氨基酸为主，长约 15～20 个氨基酸，在它的前面的一小段则以带正电荷的赖氨酸和精氨酸为特征(表 7-8)。信号序列的疏水段能够形成一段 α-螺旋结构。在信号序列之后的一段氨基酸残基也能形成一段 α-螺旋。两段 α-螺旋以反平行方式组成一个发夹结构，它很容易进入内膜的脂双层(图 7-15)。这种发夹结构在内膜中立足以后，肽链的继续合成则使信号序列以后肽链穿过内膜而进入周质。如果是内膜蛋白，则有一段或几段疏水区域永久性地镶嵌于内膜中。当组成发夹结构的两段 α-螺旋之间的部分到达内膜外表面，位于内膜外表面的信号肽酶(signal peptidase)将信号序列切除。当蛋白质全部翻译出来之后，羧端穿过内膜，在周质中折叠成蛋白质的最终构象。外膜蛋白质如何进入外膜，现在还不清楚。

翻译越膜蛋白的核糖体是与内膜结合的，这种结合至少部分地由新生肽链所介导，还需要其他什么相互作用现在不太清楚。然而，已有迹象表明，在细菌细胞内的情况很可能类似于哺乳动物细胞中的情况，有小的 RNA-蛋白质复合物参与分泌的起始和肽链延伸的控制。

### 三、真核生物蛋白质的转运、分拣和锚定

真核生物不但有细胞核、细胞质和细胞膜，而且有许多具有膜结构的细胞器。在细胞质内

## 表 7-8 细菌越膜蛋白的信号序列

| 蛋白质 | 荷电区段 | 疏水区段 | 信号肽酶作用位点↓ |
| --- | --- | --- | --- |
| **内膜蛋白：** | | | |
| 噬菌体fd 主要衣壳蛋白 | Met Lys Lys Ser Leu Val Leu Lys | Ala Ser Val Ala Val Ala Thr Leu Val Pro Met Leu Ser Phe Ala | ↓ Ala Glu Gly |
| 噬菌体fd 次要衣壳蛋白 | Met Lys Lys | Leu Leu Phe Ala Ile Pro Leu Val Val Pro Phe Tyr Ser His Ser | ↓ Ala Glu Thr |
| **周质蛋白：** | | | |
| 碱性磷酸酯酶 | Met Lys | Gln Ser Thr Ile Ala Leu Ala Leu Leu Pro Leu Leu Phe Thr Pro Val Thr Lys Ala | ↓ Arg Thr Pro |
| 麦芽糖结合蛋白 | Met Lys Ile Lys Thr Gly Ala Arg | Ile Leu Ala Leu Ser Ala Leu Thr Thr Met Met Phe Ser Ala Ser Ala Leu Ala | ↓ Lys Ile Glu |
| 亮氨酸结合蛋白 | Met Lys Ala Asn Ala Lys | Thr Ile Ile Ala Gly Met Ile Ala Leu Ala Ile Ser His Thr Ala Met Ala | ↓ Asp Asp Ile |
| β-内酰氨酶 | Met Ser Ile Gln His Phe Arg | Val Ala Leu Ile Pro Phe Phe Ala Ala Phe Cys Leu Pro Val Phe Ala | ↓ His Pro Glu |
| **外膜蛋白：** | | | |
| 脂蛋白 | Met Lys Ala Thr Lys | Leu Val Leu Gly Ala Val Ile Leu Gly Ser Thr Leu Leu Ala Gly | ↓ Cys Ser Ser |
| LamB | Met Met Ile Thr Leu Arg Lys | Leu Pro Leu Ala Val Ala Val Ala Ala Gly Val Met Ser Ala Gln Ala Met Ala | ↓ Val Asp Phe |
| OmpA | Met Lys Lys | Thr Ala Ile Ala Ile Ala Val Ala Leu Ala Gly Phe Ala Thr Val Ala Gln Ala | ↓ Ala Pro Lys |

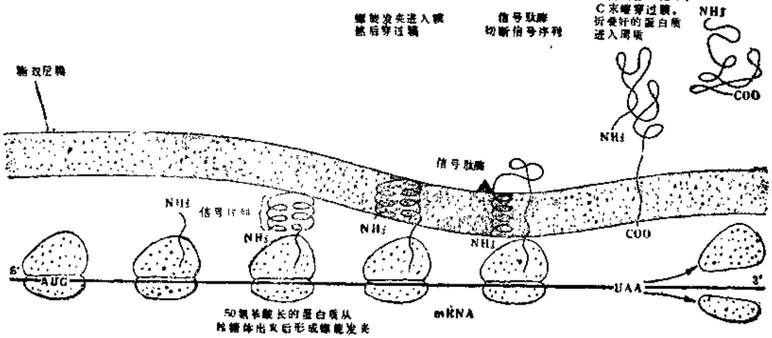



图 7-15 与翻译同时进行的分泌过程。图中示出合成分泌蛋白质的核糖体是怎样通过新生肽链与内膜结合的

合成的蛋白质是如何到达细胞的不同部位的呢？这首先要取决于蛋白质本身的结构，也就是它是否具有转运的信号。如果没有转运信号，它的命运就是留在细胞质里；而要到达细胞其他部位的蛋白质必须有转运的信号。定位于不同部位的蛋白质具有不同信号。对于核内蛋白质，线粒体内蛋白质和叶绿体内蛋白质的转运机制及其信号还不甚清楚。而较为清楚的是细胞表面蛋白质、分泌蛋白质和溶酶体蛋白质的转运途径及其信号。下面我们仅讨论这一途径。

细胞表面蛋白质、分泌蛋白质和溶酶体蛋白质(三者有时统称为分泌蛋白质)都具有与原核生物越膜蛋白相似的信号肽。这些信号肽位于 N 端，长 15～30 个氨基酸残基，大部分为疏水氨基酸，在最靠 N 端的几个氨基酸残基中，总是有带正电荷的赖氨酸或精氨酸。像原核生物的信号肽一样，能够形成包括两个 α-螺旋的发夹结构。这一结构不是插进细胞膜中，而是在信号识别颗粒(signal recognition particle, 简称 SRP)的帮助下插入到粗面内质网的膜中。

SRP 是一个小的 RNA-蛋白质复合体，它由约 300 个核苷酸的 7 SL RNA 和 6 种紧密结合的蛋白质(分子量从 9 KDa 到 72 KDa)所组成(图 7-16)。SRP 可以分为两个结构域，一个是信号识别结构域，由 7SL RNA 的大约第 100 至第 250 个核苷酸部分与 19 KDa、54 KDa、68KDa 和 72 KDa 蛋白质所构成，另一个是 7SL RNA的其余部分(相当于Alu序列，见第十一章第九节)与 9 KDa 和 14 KDa 蛋白质所构成，其功能是使肽链的延伸暂停，因而叫做延伸作用制动结构域。

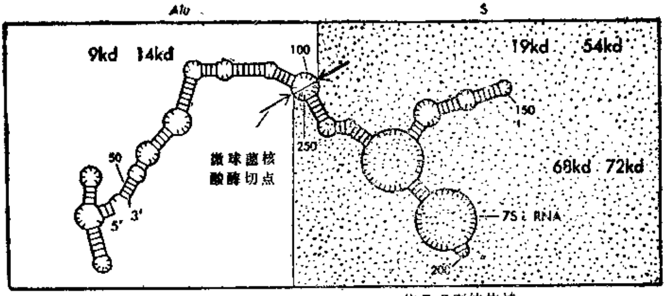



图 7-16 SRP 的两个结构域模式

SRP 的奇特性质是它能够与核糖体相结合并使肽链的延伸暂时受阻，这一作用只有在新生肽链达到大约 70 个氨基酸长时才发生，因为这时信号序列才从核糖体中冒出来，而余下的 30～40 个氨基酸仍然处于核糖体内。SRP 对翻译的阻断是暂时的，当它碰到粗面内质网膜上的 72KDa 的坞蛋白(docking protein)时，就不再阻止核糖体的翻译。这时，SRP 离开核糖体而去识别新的信号序列；而核糖体则通过信号肽与粗面内质网膜相连，并以其大亚基与膜上的核糖体受体蛋白 (ribophorin) 相互作用。尽管在分泌蛋白质的转运过程中 SRP 只起短暂的作用，但它在细胞质内的浓度很高，每个细胞内大约有 $10^6$ 个拷贝，为核糖体数目的十分之一。

SRP 作为翻译的负调控因子具有非常重要的生物学意义。许多分泌蛋白质都是降解酶类(如蛋白酶、核酸酶等)，如果有少量泄漏于细胞质内，就会产生巨大的灾难。SRP 使之合成暂停，等到为它找到安全的出路 (越膜) 后再让核糖体继续合成这些分泌蛋白质。最近有人用

RNAase 降解 SRP，得到一个大的亚复合体，具有结合核糖体并能促进分泌的能力。但这个亚复合物缺少使肽链合成暂停的功能。SRP 的两个结构域与 7SL RNA 的两种类型的序列几乎精确地相应，人们还不知其中的奥妙。

SRP 及其受体（坞蛋白）在引导分泌蛋白越膜（内质网膜）过程中的作用示于图 7-17。

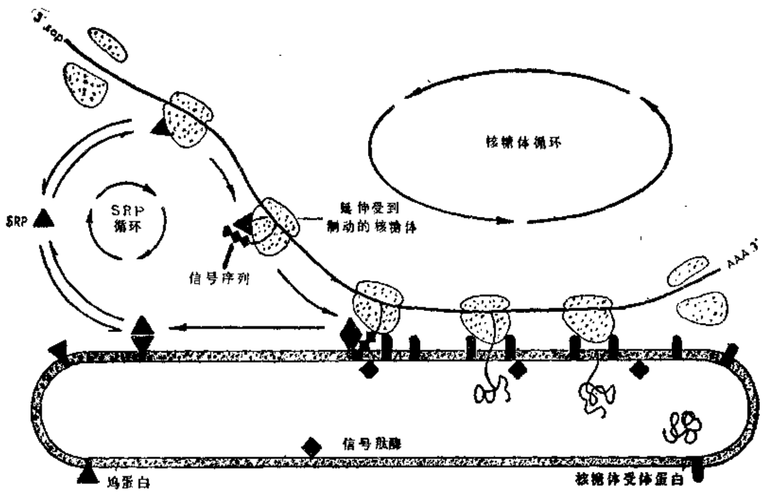



图 7-17　SRP 和 SRP 受体在引导分泌蛋白越膜过程中的作用

上述三类蛋白质进入内质网后，在内质网腔内进行肽链的折叠，并进行了初步的糖基化。然后产生转运小泡，依次进入顺面（cis）高尔基体，中间高尔基体和反面（trans）高尔基体。在反面高尔基体中完成糖基化过程并进行分拣（sorting）。分拣系统所识别的也是蛋白质分子上的信号。但这种信号不是蛋白质的一级结构，而是其空间结构的某一区域。这种信号称为信号补钉（signal patch）。不言而喻，信号补丁上的氨基酸残基在一级结构中可能相距很远。进入分泌贮备小包（secretory storage vesicles）和溶酶体的蛋白质具有这种信号补丁，而进入细胞膜和细胞表面的蛋白质可能不需要这样的信号，可能是通过一种称为“主流机制”（bulk flow）的途径由内质网经过高尔基复合体（由上述三种高尔基体组成）一直到达细胞表面（图 7-18）。这一主流机制到底是怎么一回事？其动力是什么？现在还很不清楚。此外，有些蛋白质需要滞留于某些中间环节的膜内，也需要信号补丁，当然各种膜内需要的信号补丁各不相同。

蛋白质在到达细胞膜表面（可能还有各种细胞器膜表面）是怎样固定于膜的表面的呢？人们在 70 年代初根据对人类 MN 血型蛋白（glycophorin）的研究，提出了细胞质膜的流动镶嵌模型。根据这一模型，膜蛋白分子以其疏水区插入脂双层结构，而亲水区靠近或露出膜表面，这些蛋白质至少有一个亲水区。代谢物的穿膜运输，信号的传导可能都由这些膜蛋白所介导。

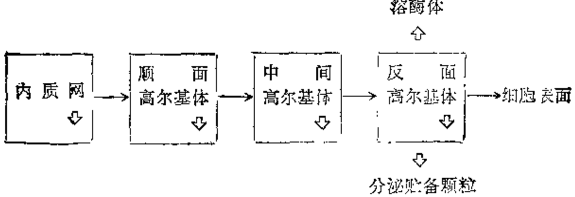



图 7-18　生物合成蛋白质的转运图解　水平箭头表示不依赖于信号的主流运输机制；竖直的空白箭头表示依赖于信号的步骤，包括进入溶酶体和分泌贮备颗粒以及各细胞器内滞留的蛋白质。

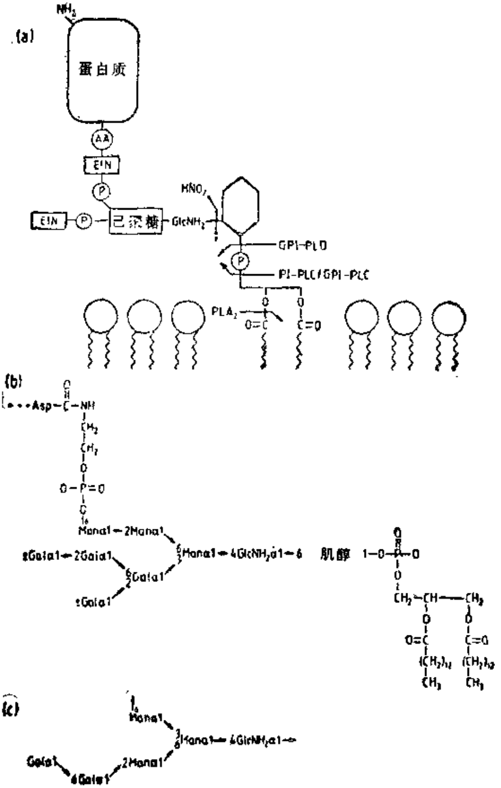



图 7-19　GPI 锚的结构　(a) GPI 结构的一般模式；(b)和(c)为锥虫的可变表面糖蛋白中己聚糖的可能连接方式。图中符号 EtN 为乙醇胺，PLC 和 PLD 分别为磷脂酶 C 和磷脂酶 D，$GlcNH_2$ 为 2-氨基葡萄糖。其他生物化学中通用符号不再注释。

然而，生物体系是极为复杂的，任何的简单模型都不能毫无例外的概括生物的某一结构功能系统。80年代中期，人们发现了鼠的胸腺细胞和神经元的 Thy-1 抗原和锥虫的可变表面糖蛋白 (VSG) 并不是通过蛋白质本身与膜结合的，而是通过一种非蛋白质锚锚定于膜的表面。这种非蛋白锚一般为脂类，共价连接于蛋白质的羧端（也有少数连接于 N 端的）。但这一羧端已经不是原初翻译产物的羧端，而是切除了原初产物羧端 15～30 个氨基酸的信号序列以后产生的羧端。这一信号序列的一致性比越膜蛋白 N 端的信号序列差，但也含较多的疏水氨基酸。这种非蛋白质锚的结构如图 7-19 所示。其主体部分为磷脂酰肌醇 (phosphatidylinositol，简称 PI)，再通过已聚糖和乙醇胺的磷酸酯与蛋白质的羧端残基以酰胺键相连。因此这类锚称为糖基-磷脂酰肌醇锚(glycosyl-phosphatidylinositol，简称 GPI)。膜表面蛋白质接 GPI 锚的机制和场所都还不清楚，估计可能发生在粗面内质网的腔面内膜上。GPI 的作用不单是将这些蛋白质固定于膜的表面，而且可能是信号传导系统的成员。因为人们最近已经发现，GPI 特异性的磷酯酶 C能切断磷脂部分的磷酸酯键，释放出二酯酰甘油；而二脂酰甘油和肌醇三磷酸酯是细胞内一对重要的第二信使。关于这方面的作用机制，人们业已着手进行深入的研究。

**四、翻译后处理(posttranslational processing)**

上面我们主要谈了与翻译同步的加工，如越膜蛋白质信号肽的去除。而 GPI 锚的连接已经属于翻译后处理工序了。这一小节继续讲解其它的翻译后处理方式。

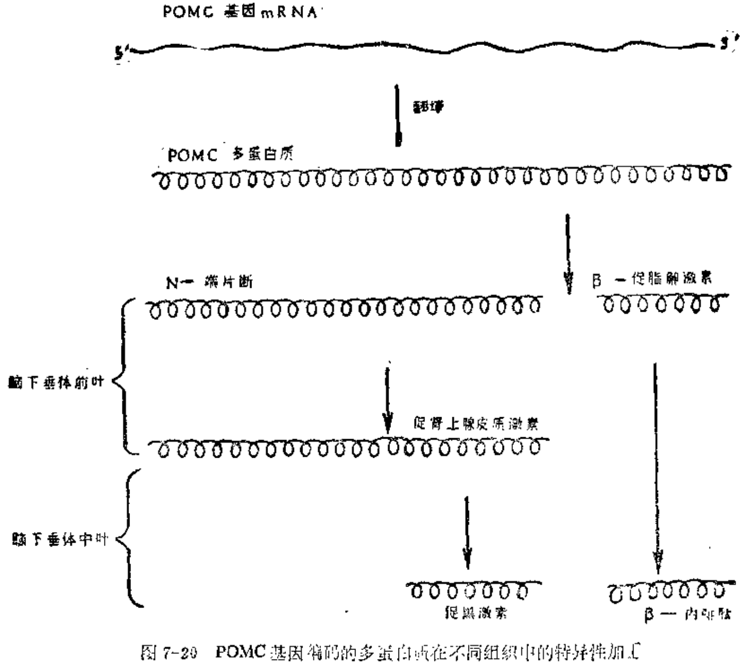



图 7-20 POMC 基因编码的多蛋白质在不同组织中的特异性加工

真核生物中像泛素基因那样翻译成多基因蛋白质的例子比较少见，但是另一种“多蛋白质”的例子倒是很多，即一种基因的翻译产物切断成为不同的蛋白质或肽。病毒蛋白质和多肽类激素的翻译后加工很普遍。哺乳动物的鸦片样促黑皮质激素原(proopiomelanocortin，简称POMC)也是一种多蛋白质，由POMC基因编码，原初翻译产物为32 KDa，然后经过复杂的加工过程(图7-26)。类似于胰蛋白酶的蛋白水解酶能够特异性地识别并切断肽链的几处序列。在脑下垂体前叶细胞中，POMC被切割成为N-端片断和C端的 $\beta$-促脂解激素。然后N-端片断又被切割成较小的N-端片断和39肽的促肾上腺皮质激素(ACTH)。而在脑下垂体中叶细胞中，在上述切割的基础上，$\beta$-促脂解激素再次被切割，产生 $\beta$-内啡肽(endorphin)；ACTH也被切割产生13肽的促黑激素($\alpha$-melanotropin)。哺乳动物脑啡肽原基因的翻译产物也要经过复杂的翻译后处理才能产生有功能的两类脑啡肽(Met 脑啡肽和 Leu 脑啡肽)。

细胞内还有一类翻译后处理，其目标不是减少氨基酸序列，而是要对已有的蛋白质产物进行共价修饰。这种共价修饰的主要反应类型有：乙酰化主要发生N末端的 $\alpha$-氨基和赖氨酸的 $\varepsilon$-氨基上；甲基化发生在 $\alpha$-氨基、$\varepsilon$-氨基、精氨酸的胍基以及C末端的 $\alpha$-羧基和侧链羧基上；磷酸化主要发生在丝氨酸的羟基上以及苏氨酸和酪氨酸的羟基上；泛素化发生在 $\alpha$-氨基和 $\varepsilon$-氨基上；转氨基酸作用发生在N-末端的 $\alpha$-氨基上；多聚ADP核糖基化主要发生在精氨酸的胍基上；糖基化可以通过N-糖苷键连接于天门冬酰胺的酰氨基上，也可以通过O-糖苷键连接于丝氨酸和苏氨酸的羟基上以及羟赖氨酸和羟脯氨酸的羟基上，也可以通过S-苷键连接于半胱氨酸的巯基上。以上每种类型的修饰作用并非都由同一种酶催化，例如 $\alpha$-氨基的甲基化与 $\alpha$-羧基的甲基化是完全不同的化学反应类型，因而催化这两种甲基化作用的酶也不相同。蛋白质的各种共价修饰作用都有极其重要的生理功能。本书限于篇幅，不能作详细的介绍。

## 提　要

DNA所荷载的蛋白质信息通过转录作用生成mRNA只完成了基因表达的一部分，更为复杂的一部分就是翻译过程，也就是核酸语言转变为蛋白质语言的过程。这一过程中的主要角色有各种氨酰基-tRNA合成酶，各种tRNA和核糖体。由于它们的共同作用，氨基酸才能按照mRNA所提供的信息相互连接成为多肽链。

所有的tRNA分子都具有相似的结构，其二级结构为三叶草形，但并不代表真实的空间构型，其空间结构为L型。许多保守的碱基在二级结构中并未形成氢键而是参与三级结构氢键的形成。除了氢键之外，相邻碱基之间堆集力也是维持其空间结构的重要因素。DNA或mRNA的四种碱基共组成64个三联体密码子，其中61个编码常规的20种氨基酸，其余3个为肽链终止密码以及少数稀有氨基酸的密码。密码子的简并性和密码子与反密码子配对的摇摆性决定了携带某种特定氨基酸的tRNA种类的最低数目；但实际上存在的tRNA种类还要多一些。其中少数tRNA基因是抑制tRNA的主要来源。一组同功tRNA由同一种氨酰基-tRNA合成酶催化而携带氨基酸。对于大多数tRNA来说，氨酰基-tRNA合成酶并非识别反密码子，而是识别其它某个部位，如丙氨酸抑制tRNA的接受臂上的G3:U70，这就是tRNA的副密码子。密码子的使用频率与细胞内识别密码子的tRNA含量呈正相关，特别是需要量大的蛋白质更是如此。适中的密码子-反密码子作用强度有利于翻译的迅速进行。从原核生物直至最高等的真核生物，都使用相同的密码子，这就是密码子的通用性。然而，后来人们发现线粒体密码子却有若干不同之处。

大肠杆菌核糖体的沉降系数为70S，包括55个蛋白质分子和3个rRNA分子，其中大亚基有34个蛋白质和23S rRNA及5S rRNA，小亚基有21个蛋白质和16S rRNA。核糖体和其他辅助因子在一起提供了翻译过程的全部酶活性，这些酶活性只有在核糖体整体结构完整的情况下才会具备。真核生物的核糖体则更为复杂。

在原核生物中，翻译的起始涉及到起始复合物和70S核糖体的形成。这个过程需要带有S-D序列的mRNA，三种起始因子，fMet-tRNA$_f$，30S亚基和50S亚基。真核生物的起始复合物并不是在起始密码子处形成，而是在帽子结构处形成，然后才移动到AUG处与60S亚基结合成为80S核糖体。真核生物的起始因子则多达10个以上，其起始机制亦更为复杂。延伸过程则包括转肽与肽键的形成，转位以及Ts循环等十分复杂的步骤。翻译的终止和肽链的释放是在核糖体，终止密码子和释放因子共同作用之下完成的。原核生物有三种释放因子，真核生物只有一种释放因子。

mRNA的结构与翻译作用密切相关。除了mRNA上的核糖体识别位点(原核生物的S-D序列和真核生物的帽子结构)，起始密码子和终止密码子是翻译所必需之外，还有许多结构特点与翻译有关。例如，原核生物的多基因mRNA的基因间间隔区的长短以及是否有自己的S-D序列均影响后续基因的翻译效率。真核生物中有时能选择不同的AUG作为起始密码子。5′端非编码区域对翻译也具有调控作用。真核生物中尾首相连的多聚泛素基因的翻译则需要特异的酶进行翻译后处理。

严格说来，翻译的完成还不是基因表达的完成，合成的蛋白质还必须有条不紊地到达各自的岗位。这就是细胞内蛋白质的越膜运输的问题。在这个过程中，信号肽序列起重要的作用。真核细胞的结构要比原核细胞复杂得多，因而越膜运输的机制也要复杂得多。在越膜过程中，有时还有共翻译处理过程以及许多复杂的翻译后处理过程。

## 参 考 文 献


1. Rich, A. and RajBhandary, U. L. (1976) "Transfer RNA: Molecular Structure, Sequence, and Properties." *Ann. Rev. Biochem.* 45: 805～860.
2. Kim, S. H. et al. (1974) "Three-dimensional Tertiary Structure of Yeast Phenylalanine Transfer RNA." *Science* 185: 435～440.
3. Robertus, J. D. et al. (1974) "Structure of Yeast Phenylalanine tRNA at 3 Å Resolution." *Nature* 250: 546～551.
4. Rich, A. and Kim, S. H. (1978) "The Three-dimensional Structure of Transfer RNA." *Sci. Amer.* 238: 56～62.
5. Lagerkvist, U. (1981) "Unorthodox Codon Reading and the Evolution of the Genetic Code." *Cell* 23: 305～306.
6. Crick, F. H. C. (1966) "Codon-anticodon Pairing: the Wobble Hypothesis." *J. Mol. Biol.* 19: 548～555.
7. Normanly, J. et al. (1986) "Changing the Identity of a Transfer RNA." *Nature* 321: 213～219.
8. Schimmel, P. (1987) "Aminoacyl tRNA Synthetases: General Scheme of Structure-function Relationships in the Polypeptides and Recognition of Transfer RNAs." *Ann. Rev. Biochem.* 56: 125～158.
9. Hou, Y.-M. and Schimmel, P. (1988) "A Simple Structural Feature is a Major Determinant of the Identity of a Transfer RNA." *Nature* 333: 140～145.
10. de Duve, C. (1988) "The Second Genetic Code." *Nature* 333: 117～118.
11. Grantham, R. (1981) "Codon Catalog Usage Is a Genome Strategy Modulated for Gene Expressivity." *Nucl. Acid Res.* 9: 943～974.
12. Ikemura, T. and Ozeki, H. (1982) "Codon Usage and Transfer RNA Contents, Organism-specific Codon-choice Patterns in Reference to the Isoacceptor Contents" *Cold Spring Harbor Symp. Quant. Biol.* 47: 1087～1096.

13. Kurland, C. G. (1987) "Strategies for Efficiency and Accuracy in Gene Expression. 1. The Major Codon Preference: a Growth Optimization Strategy." *Trends in Biochem. Sci.* 12: 126～128.
14. Grosjean, F. and Fiers, W. (1982) "Preferential Codon Usage in Prokaryotic Genes: The Optimal Codon-anticodon Interaction Energy and the Selective Codon Usage in Efficiently Expressed Genes." *Gene* 18: 199～209.
15. Anderson, S. et al. (1981) "Sequence and Organization of the Human Mitochondrial Genome." *Nature* 290: 457～465.
16. Attardi, G. (1985) "Animal Mitochondrial DNA: an Extreme Example of Genetic Economy." *Internatl. Rev. Cytol.* 93: 93～145.
17. Engelberg-Kulka, H. and Schoulaker-Schwarz, R. (1988) "A Flexible Genetic Code, or Why Does Selenocysteine Have No Unique Codon?" *Trends in Biochem. Sci* 13: 419～421.
18. Leinfelder, W. et al. (1988) "Gene for a Novel tRNA Species that Accepts L-Serine and Cotranslationally Inserts Selenocysteine." *Nature* 331: 723～725.
19. Shepherd, J. C. W. (1982) "From Primeval Message to Present-day Gene." *Cold Spring Harbor Symp. Quant. Biol.* 47: 1099～1108.
20. Crothers, D. M. (1982) "Nucleic Acid Aggregation Geometry and the Possible Evolutionary Origin of Ribosomes and the Genetic Code." *J. Mol. Biol.* 162: 379～391.
21. Wittmann, H. G. (1982) "Components of Bacterial Ribosomes." *Ann. Rev. Biochem.* 51: 155～183.
22. Wittmann, H. G. (1983) "Architecture of Prokaryotic Ribosomes." *Ann. Rev. Biochem.* 52: 35～65.
23. Noller, H. F. (1984) "Structure of Ribosomal RNA." *Ann. Rev. Biochem.* 53: 119～162.
24. Nierhaus, K. H. (1982) "Structure, Assembly, and Function of Ribosomes." *Current Topics in Microbiol. and Immun.* 97: 82～155.
25. Rohl, R., and Nierhaus, K. H. (1982) "Assembly Map of the Large Subunit (50S) of *Escherichia coli* Ribosomes." *Proc. Nat. Acad. Sci.* 79: 729～733.
26. Finley, D. et al. (1989) "The Tails of Ubiquitin Precursors Are Ribosomal Proteins Whose Fusion to Ubiquitin Facilitates Ribosome Biogenesis." *Nature* 338: 394～401.
27. Redman, K. L. and Rechsteiner, M. (1989) "Identification of the Long Ubiquitin Extension as Ribosomal Protein S27a." *Nature* 338: 438～440.
28. Kozak, M. (1983) "Comparison of Initiation of Protein Synthesis in Prokaryotes, Eukaryotes, and Organelles." *Microbiol. Rev.* 47: 1～45.
29. Shine, J., and Dalgarno, L. (1974) "The 3′ Terminal Sequence of *E. coli* 16S rRNA: Complementarity to Nonsense Triplets and Ribosome Binding Sites." *Proc. Nat. Acad. Sci.* 71: 1342～1346.
30. Gold, L. et al. (1981) "Translational Initiation in Prokaryotes." *Ann. Rev. Microbiol.* 35: 365～403.
31. Gold, L. (1988) "Posttranscriptional Regulatory Mechanisms in *Escherichia coli*." *Ann. Rev. Biochem.* 57: 199～233.
32. Gupta, N. K. et al. (1987) "Roles of Eukaryotic Initiation Factor 2 and Eukaryotic Initiation Factor 2 Ancillary Protein Factors in Eukaryotic Protein Synthesis Initiation." In *Translational Regulation of Gene Expression*, Ed. by J. Ilan (Plenum Press) pp. 287～334.
33. Merrick, W. C. et al. (1987) "Involvement of Nucleotides in Protein Synthesis Initiation." In *Translational Regulation of Gene Expression*, Ed. by J. Ilan (Plenum Press) pp. 265～286.
34. Garrett, R. A., and Woolley, P. (1982) "Identifying the Peptidyl Transferase Centre." *Trends in Biochem. Sci.* 7: 385～386.
35. Gehrke, L., and Ilan, J. (1987) "Regulation of Messenger RNA Translation at the Elongation Step During Estradiol-induced Vitellogenin Synthesis in avian liver." In *Translational Regulation of Gene Expression*. Ed. by J. Ilan (Plenum Press), pp. 165～185.
36. Johnson, A. et al. (1982) "Distance Moved by Transfer RNA During Translocation from the A Site to the P Site on the Ribosome." *J. Mol. Biol.* 156: 113～140.
37. Moldave, K. (1985) "Eukaryotic Protein Synthesis." *Ann. Rev. Biochem.* 54: 1109～1119.
38. Shatkin, A. J. (1985) "mRNA Cap Binding Proteins: Essential Factors for Initiating Translation." *Cell* 40: 223～224.
39. Laz, T. et al. (1987) "The Role of Messenger RNA Sequences and Structures in Eukaryotic Translation." In

*Translational Regulation of Gene Expression*, Ed. by J. Ilan (Plenum Press), pp. 413～429.
40. Özkaynak, E. et al. (1937) "The Yeast Ubiquitin Genes: a Family of Natural Gene Fusions." *The EMBO Journal* 6: 1429～1439.
41. Engelman, D. M., and Steitz, T. A. (1981) "The Spontaneous Insertion of Proteins into and Across Membranes: the Helical Hairpin Hypothesis." *Cell* 23: 411～422.
42. Michaelis, S., and Beckwith, J. (1982) "Mechanism of Incorporation of Cell Envelope Proteins in *Escherichia coli*." *Ann. Rev. Microbiol.* 36: 435～465.
43. Davis, N. G., and Model, P. (1985) "An Artificial Anchor Domain: Hydrophobicity Suffices to Stop Transfer." *Cell* 41: 607～614.
44. Silhavy, T. J. et al. (1983) "Mechanisms of Protein Localization." *Microbiol Rev.* 47: 313～344.
45. Meyer, D. I. et al. (1982) "Secretory Protein Translocation across Membranes —— the Role of the Docking Protein." *Nature* 297: 647～650.
46. Walter, P. et al. (1984) "Protein Translocation across the Endoplasmic Reticulum." *Cell* 38: 5～8.
47. Pfeffer, S. R., and Rothman, J. E. (1987) "Biosynthetic Protein Transport and Sorting by the Endoplasmic Reticulum and Golgi." *Ann. Rev. Biochem.* 56: 829～852.
48. Ferguson, M. A. J., and Williams, A. F. (1988) "Cell-Surface Anchoring of Proteins Via Glycosyl-Phosphatidylinositol Structures." *Ann. Rev. Biochem.* 57: 285～320.
49. Berridge, M. J. (1987) "Inositol Triphosphate and Diacylglycerol: Two Interacting Second Messengers." *Ann. Rev. Biochem.* 56: 159～193.
50. Douglass, J. et al. (1984) "Polyprotein Gene Expression: Generation of Diversity of Neuroendocrine Peptides." *Ann. Rev. Biochem.* 53: 665～715.
51. Mains, R. E. et al. (1983) "Strategies for the Biosynthesis of Bioactive Peptides." *Trends in Neurosci.* 6: 229～235.
52. Hirschberg, C. B. and Snider, M. D. (1987) "Topography of Glycosylation in the Rough Endoplasmic Reticulum and Golgi Apparatus." *Ann. Rev. Biochem* 56: 63～87.
53. Towler, D. A. et al. (1988) "The Biology and Enzymology of Eukaryotic Protein Acylation." *Ann. Rev. Biochem.* 57: 69～99.

# 第八章 原核生物基因表达的调控

## 第一节 概　　述[1]

原核生物是单细胞生物，没有核膜和明显的核结构。它与周围环境密切相关，它本身既无足够的能源贮备，又无高等植物那种制造有机物的本领。原核生物在长期进化过程中演化出来的适应性和高度的应变能力，是它们赖以生存繁衍的根本。细菌的生长与周围环境密切相关，它们必须不断调节各种不同的基因的表达以适应营养条件(如碳源，氮源等)和对付周围不利的物理化学因素(如高温，射线，重金属，烷化剂等)。细菌必须能迅速合成自身需要的蛋白质(酶)、核酸和其他生物大分子，而同时又能迅速地停止合成和降解那些不再需要的成分。原核生物的细胞结构以及基因表达的调控方式都是与这些要求相适应的。即使是最简单的噬菌体，其中早期基因和晚期基因的表达也是严格调控的，否则就不能有效地进行其生长和繁殖。例如噬菌体的溶菌酶基因的表达就是严格控制的，只有当噬菌体合成了大量的 DNA 和外壳蛋白，才使细菌裂解。否则，噬菌体就不能产生大量的后代。

原核生物与真核生物一样，其基因调控主要是发生在转录水平上。其中的道理也很简单。任何一系列连锁过程，控制其第一步是最有效的和最经济的。原核生物中的操纵元系统就是这种最有效和最经济原则的体现。因为把生物活性相关的基因组织在一起，就不必逐个地进行调控，而是一开俱开，一关全关，同时还保持基因产物在比例上大体相当，甚至是精确的比例(如 rRNA 基因，见第二章)。转录水平的调控形式除了操纵元之外，还有其他各种形式，如噬菌体基因表达的时序调控。有时原核生物也从 DNA 水平对基因表达进行调控，如沙门氏菌的相变过程，但其实质是以基因重排的方式控制基因的转录。还有以翻译方式来控制基因转录的，如色氨酸操纵元的衰减子系统。

在原核生物中，也有不少控制翻译过程的调控机制。RNA 噬菌体 mRNA 的二级结构对基因表达的调控，大肠杆菌外膜蛋白合成过程中反义 RNA 的调控作用，λ 噬菌体整合酶合成过程中通过影响 mRNA 的稳定性而进行的调控等，都属于这一范畴。

应当指出，转录的起始，延伸，终止和抗终止过程及翻译的起始，延伸和终止过程，每一步都在对基因的表达实行调控。然而，我们已将其中不少内容放在第六章和第七章中讲授，因为这些内容也是转录和翻译过程的基本方面。

## 第二节 正调控和负调控[1~3]

严格说来，正调控应该称为正控制 (positive control)；同样，负调控应该称为负控制 (negative control)。正控制系统和负控制系统是按照没有调节蛋白质存在的情况下操纵元对于新加入的调节蛋白质的响应情况来定义的。在没有调节蛋白质存在时基因是表达的，如入这种调节蛋白后基因表达活性便被关闭。这样的控制系统就叫做负控制系统。负控制系统中的调节蛋白质称为阻遏蛋白(repressor)。相反，如果在没有调节蛋白质存在时基因是关闭的，

加入这种调节蛋白质后基因活性就被开启。这样的控制系统就叫做正控制系统。正控制系统中调节蛋白质叫做无辅基诱导蛋白(apoinducer)。这种所谓开启和关闭的调控主要是允许基因转录产生 RNA 还是防止基因转录产生 RNA；也有通过影响核糖体对 mRNA 翻译的效率而控制酶活性的。这里所说的“关闭”并不是绝对的，实际上并不存在使一个操纵元彻底关闭的例子。当我们说一个操纵元处于关闭状态，意思是说它的表达很低；但是总还残留着本底水平的基因表达。通常在一个细胞周期中，这样的表达只能产生 1～2 个 mRNA 分子。甚至是所谓开启，也常有程度的差异。

负控制提供了一个万一保安机制，即万一调节蛋白质失活，酶系统可照样合成，只不过有时浪费一点而已；而决不会使细胞因缺乏该酶系统而造成致命的后果。因此我们容易设想这种控制系统是如何进化的。起初，这种酶系统的表达是组成型的，后来细胞中产生一种干预这种表达的机制而给细胞提供了选择优势，这种机制就演化成现在的负控制系统。而对于正控制系统，就很难推演其进化过程。但可以肯定的一点是，细胞在获得任何调控机制之前就应该具有使这些基因表达的能力。很可能这种正控制因子原来是用作基因表达的常规成分，而后来只在一部分基因的表达过程中保留下来作为正控制因子。

根据操纵元对于能调节它们表达的小分子的应答反应的性质，可将操纵元分为可诱导的操纵元(inducible operon)和可阻遏的操纵元(repressible operon) 两大类。在可诱导的操纵元中，加入这种对基因表达有调节作用的小分子后，则开启基因的转录活性。这种作用及其过程叫做诱导 (induction)，产生诱导作用的小分子物质叫做诱导物 (inducer)。在可阻

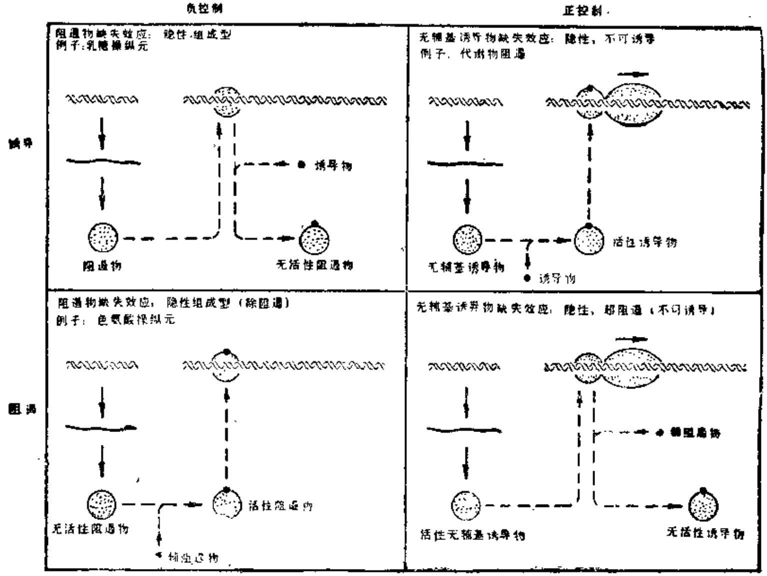



图 3-1 灵活可变的控制网络使正控制系统和负控制系统都能进入诱导状态或阻遏状态

遏的操纵元中，加入对基因表达有调节作用的小分子物质后，则关闭基因的转录活性。这种作用及其过程叫做阻遏(repression)，产生阻遏作用的小分子物质叫做辅阻遏物(corepressor)。可诱导操纵元处于转录活动状态时，我们说该操纵元已被诱导(induced)；可阻遏的操纵元处于转录活动状态时，我们说该操纵元已除阻遏(derepressed)，有时亦称被诱导(induced)。在某些条件下，如调节蛋白质突变，使可诱导的操纵元变为不可诱导(uninducible)；某些条件则使可阻遏操纵元变为不能除阻遏，则称为超阻遏(superrepressed)。对于可诱导的操纵元，这种状况有时也称为超阻遏。

无论是正控制还是负控制，都可以通过调节蛋白质与小分子物质(诱导物或辅阻遏物)的相互作用而达到诱导状态或阻遏状态。图 8-1 总结了四种简单类型的控制网络。当诱导物钝化了阻遏蛋白或活化了无辅基诱导蛋白，操纵元就达到了诱导状态。当辅阻遏物活化了无活性的阻遏蛋白(有时亦称无辅基阻遏蛋白，aporepressor)或者钝化了无辅基诱导蛋白，则操纵元进入受阻遏状态。在遗传学上利用调节蛋白基因的突变(产生无功能的蛋白质)所产生的后果来区分负控制系统和正控制系统。阻遏蛋白的失活产生隐性的组成型表达，这是负控制系统的特征；无辅基诱导蛋白的失活导致不可诱导状态或超阻遏状态，则是正控制系统的特征。

## 第三节　操纵元的原型——乳糖操纵元[1,4~7]

### 一、操纵元模型的提出

人们早在 20 世纪初就发现了酵母细胞中酶的诱导现象，即分解底物的酶只有在底物存在时才出现。酶的可诱导性在细菌中普遍存在，人们对大肠杆菌的酶诱导现象研究最为透彻。当大肠杆菌生长在没有乳糖的培养基上时，每个细胞内只有不到 5 个分子的 β-半乳糖苷酶(负责将乳糖分解成为其构成组分即半乳糖和葡萄糖这两个单糖，然后再进一步代谢)。当加入适合的底物如乳糖或其他半乳糖苷后，在 2～3 分钟内开始出现 β-半乳糖苷酶，很快可以达到每个细胞 5000 个酶分子。在适合的条件下，β-半乳糖苷酶有时可占细胞蛋白总量的 5～10%。如果撤除底物，则 β-半乳糖苷酶的合成迅速停止，就像当初加入底物时迅速合成一样。

细菌细胞中的酶有的可以被诱导，有的则可以被阻遏。例如 *E. coli* 的负责色氨酸合成的色氨酸合成酶就是明显的例子。当培养基中没有色氨酸时，该基因表达；而加入色氨酸后则合成迅速停止。这种效应就叫做阻遏效应。

由上可见，细菌一方面能调节它分解底物的能力，另一方面又能调节它合成某种物质的能力。这两种过程的调节者都是小分子物质。如果某种物质能够促使细菌产生酶来分解它，这种物质就是诱导物(inducer)；如果某种物质能够阻止细菌产生合成这种物质的酶，这种物质就是辅阻遏物(corepressor)。

对于 β-半乳糖苷酶来说，除了能被该酶分解的半乳糖苷(如乳糖)可以作为诱导物外，一些能被该酶识别但不能被分解的半乳糖苷化合物也能作为诱导物，如异丙基-β-D-硫代半乳糖苷(isopropyl-β-D-thiogalactoside，简称 IPTG)。IPTG 有极强的诱导效应，而本身又不被分解，因而被称为安慰诱导物(gratuitous inducer)。由于 IPTG 在诱导过程中不被分解，因而给试验分析带来极大的方便。IPTG 的结构如图 8-2 所示。

在细菌中，同一代谢途径的所有的酶的合成调控大多是一致的。大肠杆菌中分解利用乳糖的酶除了β-半乳糖苷酶之外，还需要使乳糖能够进入细胞的酶即β-半乳糖苷透性酶以及β-半乳糖苷乙酰基转移酶，它能将乙酰基从乙酰辅酶A分子上转移到β-半乳糖苷上，其生物学意义尚不完全清楚。这三个基因(Z、Y、A)由同一个启动子控制，转录成一条多基因的mRNA，然后依次翻译出β-半乳糖苷酶，透性酶和转乙酰酶。*lac ZYA*的转录是如何调控的呢？法国分子生物学家Jacob和Monod等人分离了大量的*lacZYA*系统的突变体，其中有的突变体能在没有诱导物存在的情况下合成Lac酶系统。这就是我们在第五章中所讲的组成型突变。遗传学分析证明，这些组成型突变的突变位点集中在两个"基因"位点上，即*lacI*和*lacO*。这两个位点对*lacZYA*调控的机制是什么呢？Jacob和Monod将这些突变"基因"分别与大肠杆菌的性因子即F因子相连成为F′因子，然后将这些F′因子再导入大肠杆菌细胞中，就产生部分二倍体(partial diploid)以便进行互补试验等遗传学分析。兹将Jacob和Monod以及其他人的部分实验结果选列于表8-1。

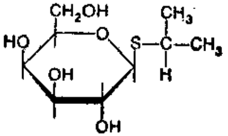



图8-2 IPTG的分子结构

**表8-1 大肠杆菌 *lac* 操纵元部分二倍体基因表达情况**

| 组　别 | 基　　因　　型 | *lac*基因的表达：C表示组成型，I表示诱导型 |
|---|---|---|
| 1 | $lacI^+O^+Z^+/F'lacI^+O^cZ^+$ | C |
| 2 | $lacI^+O^cZ^+/F'lacI^+O^+Z^+$ | C |
| 3 | $lacI^+O^+Z^-/F'lacI^-O^+Z^+$ | I |
| 4 | $lacI^-O^+Z^+/F'lacI^+O^+Z^-$ | I |
| 5 | $lacI^+O^+Z^+/F'lacI^+O^cZ^-$ | I |
| 6 | $lacI^-O^+Z^+Y^-/F'lacI^+O^cZ^-Y^+$ | Y:C<br>Z:I |

在表8-1中，第1组和第2组的实验结果表明，$lacO^c$对于$lacO^+$来说为显性，而第1组和第5组则说明$lacO^c$的显性作用只有在它和$lacZ^+$处于同一个DNA分子上(其实还必须在特定位置)才能表现出来。这就说明*lacO*的作用方式为顺式作用。第3组和第4组实验说明了$lacI^+$的作用方式为反式作用，即通过产生可扩散的物质(RNA或蛋白质，*lacI*产物为蛋白质)来发生作用。综上所述，*lacO*位点的作用为顺式作用，而*lacI*基因的作用为反式作用。第6组实验则同时说明了这两种成分的作用方式。

除了组成型突变体以外还有另一类突变体，在有无诱导物时都不能合成Lac酶系统(结构基因正常)。Jacob和Monod根据突变点的遗传学位置，把这类突变又分为两个亚类：一种是$lacI^s$突变，对野生型$lacI^+$为反式显性；另一种位于启动子*lacP*中，其作用方式为顺式作用(当时人们还不知道启动子的概念)。

根据以上结果，Jacob和Monod于1961年提出了操纵元模型。其要点是：一个或几个结构基因与一个调节基因和一个操纵位点组成一个操纵单元。这个单元称为操纵元(operon)，

这个操纵位点就称为操作子(operator)。在操纵元中，结构基因产生 mRNA 并作为模板合成蛋白质；而调节基因则产生一种阻遏蛋白与操纵位点相互作用；这个操纵位点总是与它所控制的结构基因相毗邻(注：后来证明，操纵位点并非都与结构基因相毗邻！)；阻遏蛋白与操纵位点结合从而阻碍了结构基因转录成为 mRNA；在诱导过程中，诱导物(又称为抗阻遏物 antirepressor)通过与阻遏蛋白相结合而阻止阻遏蛋白与操纵位点的结合。1964 年，人们发现了启动子。因此启动子也就成为操纵元的一个成分。这样，就得到了我们在第六章中早已阐述过的关于操纵元的完整的概念。

## 二、乳糖操纵元的调控机理

*lac* 操纵元的基本调控途径示于图 8-3。*lacZ*，*lacY* *lacA* 为结构基因，溯流而上依次为操作子(操纵位点)，启动子和调节基因 *lacI*。当细胞内无诱导物存在时，阻遏蛋白能够与操作子结合。由于操作子与启动子有一定程度的重叠，因此妨碍了 RNA 聚合酶在-10序列上形成

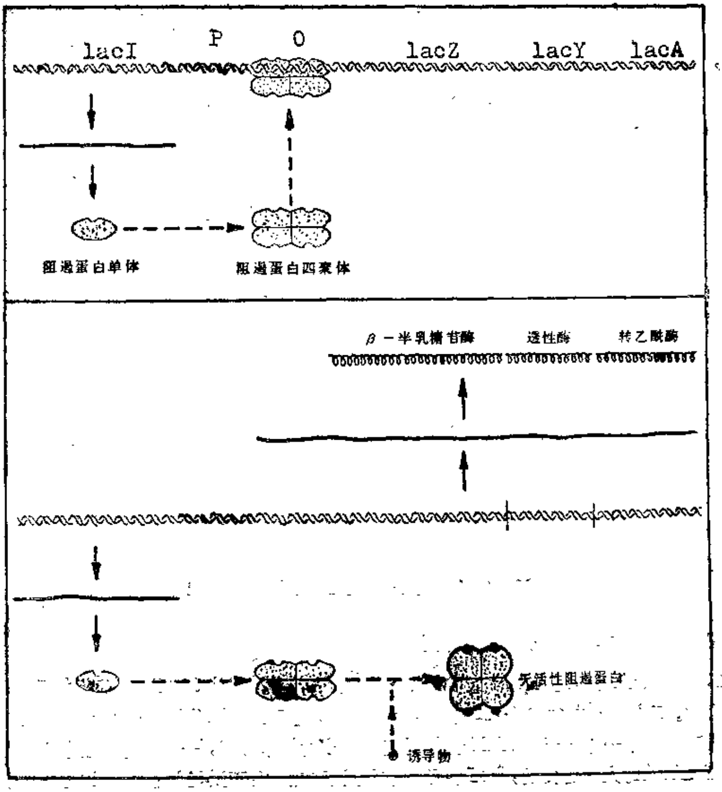



图 8-3 *lac* 操纵元被诱导物开启

开放性启动子复合物。当细胞内有诱导物存在时，诱导物以其极高的浓度与阻遏蛋白迅速结合，从而改变阻遏蛋白的构象，使之迅速地从操作子上解离下来。这样，RNA 聚合酶就能与启动子牢固结合并形成开放性启动子复合物，从而开始转录 *lacZYA* 结构基因。这种通过与小分子结合而进行的变构控制在基因转录的调控中是很普遍的现象。关于 *lac* 阻遏蛋白的变构机制将在下一小节中讨论。

另外有几个问题还需要说明一下。首先，在非诱导状态下乳糖操纵元(以及其他许多操纵元)存在着本底组成型合成(background constitutive synthesis)。这是因为阻遏蛋白对于 *lacO* 的结合并非是经久不变的，而是有一个 10～20 分钟的半衰期。虽然 *lacO* 处于游离状态的时间极短，但有时 RNA 聚合酶就抓住这一时机进行一次转录。平均每个细胞周期有一个 *lac* mRNA 以此方式产生，从而翻译出几个分子的β-半乳糖苷酶，β-半乳糖苷透性酶和β-半乳糖苷转乙酰酶。也就是说有几个β-半乳糖苷透性酶在巡逻值班，一旦培养基中有β-半乳糖苷物质存在时，就能允许它们进入细胞成为诱导物。这就是最初的诱导物为什么能进入细胞的道理。其次，后来人们发现，原来被认为是诱导物的乳糖实际上并不能与阻遏蛋白结合。那么乳糖又是怎样诱导乳糖操纵元基因的表达的呢？原来，乳糖进入细胞后由β-半乳糖苷酶催化，变为别乳糖(allolactose)，别乳糖是一种较强的诱导物。这一转变过程示于图 8-4。当细胞内的乳糖和别乳糖被代谢完毕后，阻遏蛋白又可以结合于操作子，*lac* 操纵元的结构基因又处于关闭状态。

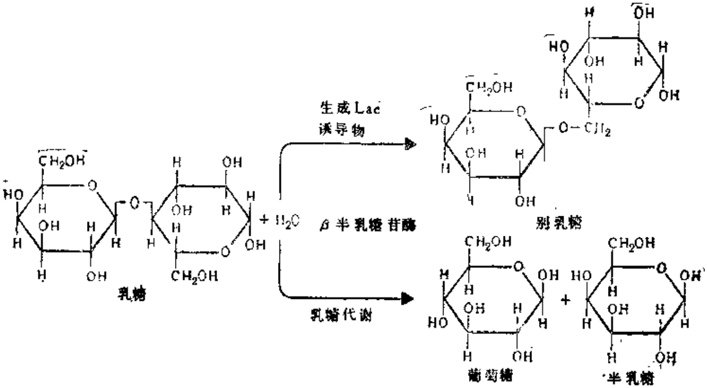



图 8-4 β-半乳糖苷酶能催化乳糖的两种反应

### 三、阻遏蛋白与操作子的相互作用

1. 操作子上与阻遏蛋白结合的位点　阻遏蛋白的分子量为 38 KDa，四个相同的亚基组成四聚体。每个细胞中含有 5～10 个这样的四聚体。这可以用 IPTG 结合试验测定出来。每个亚基都可以结合一个分子的 IPTG。*lacI* 基因的表达效率很低，一个细胞周期中只合成 1～2 个 *lacI* mRNA。人们还发现，*lacI* 基因的转录是组成型的，不受任何阻遏蛋白控制，而只受自身启动子强度的控制。由于 *lacI* 启动子与 RNA 聚合酶的亲和性很低，因此转录水平也很低。

目前已获得 *lacI* 的启动子上升突变体，使细胞中 *lac* 阻遏蛋白含量剧增，有时能达到细胞蛋白总量的1%。由于细胞中阻遏蛋白的浓度很高，而少量的诱导物分子不能完全中和它，因而这些突变体的 Lac 酶系统是难于诱导的。人们将这些突变体的细胞破碎，然后与 IPTG 结合，从而分离和纯化 *lac* 阻遏蛋白。

人们最初用硝酸纤维膜鉴定阻遏蛋白与 DNA 的结合。用野生型的 DNA 片断与阻遏蛋白就产生能结合于膜的复合物，并能分离出这些与阻遏蛋白结合的 DNA 片断。这种结合可以为 IPTG 所抑制。而用 *lacO*$^c$ 突变体的 DNA 片断，则不能与阻遏蛋白相结合。这就说明了乳糖操纵元模型中阻遏蛋白确实是与操作子相结合的。关于阻遏蛋白在操作子上的结合位点的分析，其方法与 RNA 聚合酶结合位点的分析方法相同，这里就不再重复(见第六章)。用这些方法分析出的阻遏蛋白保护区域为 26 bp(－5～＋21)。这一区域及其中的结合位点示于图 8-5。

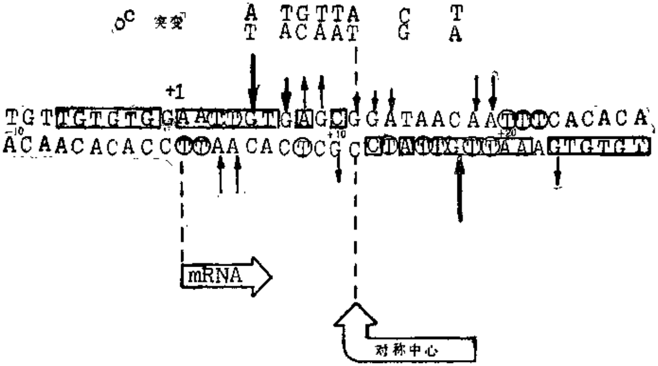



图 8-5 lac 操作子的对称序列。阴影部分表示反向重复序列。加圈的 T 在用5-溴尿嘧啶取代后可与阻遏蛋白交联。指向 DNA 的箭头表示可被阻遏蛋白保护而免遭甲基化的碱基；背离 DNA的箭头表示甲基化被加强了的碱基。箭头长度表示效应的强弱。

阻遏蛋白结合位点的 DNA 序列的对称性与阻遏蛋白四聚体的对称性是相应的。以＋11为对称轴，每边为两段各 6 个 bp 的对称序列(TGTGTG 和 AATTGT)。从图 8-5 可以看出，靠近对称轴的一对反向重复序列在与阻遏蛋白结合过程中起主要作用。甲基化试验表明，＋3～＋19 区段中的许多嘌呤碱基(A 和 G)可被阻遏蛋白保护而免于甲基化；相反，由于阻遏蛋白与 *lacO* 的结合而提高了少数嘌呤碱基对甲基化的敏感程度。这可能是由于在 DNA-阻遏蛋白的复合物中形成了一个憎水的口袋，从而加强了甲基化作用。而紫外线交联实验表明，位于＋1～＋22 之间的绝大多数胸腺嘧啶在用 5-溴尿嘧啶取代后都能与阻遏蛋白形成交联，说明这些 T 都与阻遏蛋白紧密相连(见图 8-5)

根据组成型突变体突变位点的分析，从＋5到 ＋17 这一狭小的范围(13 bp)内的核苷酸序列更为重要，因为现已发现的组成型突变全部集中在这个范围内，其中有 8 个碱基对的突变能造成 *lac* 基因的组成型表达。这些组成型突变位点，大部分位于对称轴的左侧，说明左侧序列只有保持野生型结构才能与阻遏蛋白牢固结合以阻碍 RNA 聚合酶在－10 序列附近的牢固结

合；而右侧序列的改变虽然影响到阻遏蛋白结合的牢固程度，但由于左侧的牢固结合仍能有效地阻碍 RNA 聚合酶在 −10 序列附近的牢固结合。实际上，左侧序列对于阻遏蛋白的亲和性更高。人工合成的仅由左侧序列(−7～+10)组成二倍旋转对的操作子 (34 bp 的反向重复序列)，它与阻遏蛋白的亲和性比天然的乳糖操作子高 10 倍，这主要可能是由于右侧序列的转变而造成的。

根据对 RNA 聚合酶结合位点和阻遏蛋白结合位点的分析，二者的主要结合位点并不相互重叠而是相邻。由于这两种蛋白质分子都很大，因而阻遏蛋白的结合可能妨碍RNA 聚合酶在 −10 序列上的结合。然而，新近的研究表明，Lac 阻遏蛋白并不是阻碍 RNA 聚合酶的结合，而可能是阻遏了 RNA 聚合酶-启动子复合物的异构化。

2. 阻遏蛋白上与操作子结合的位点　当用胰蛋白酶对阻遏蛋白进行切割时，可以从第59和第 60 氨基酸残基之间切断。从第 60 氨基酸残基直到 C 末端的第 360 氨基酸称为“抗胰蛋白酶核心”，它保留了形成四聚体以及与诱导物结合的能力，但是它不能和操作子结合。而N端的第 1～59 氨基酸残基组成的片断称为长头片断 (long headpiece)，长头片断又可以被胰蛋白酶在第 51、第 52 氨基酸残基之间切断。N 端 51 个氨基酸残基的片断则称为短头片断 (short headpiece)。长头片断和短头片断都保留与操作子结合的能力，而且与 *lacO* 的接触方式与完整的阻遏蛋白完全一样。N 末端的 50 个氨基酸残基像手臂一样从核心片断上伸出，负责与操作子 *lacO* 结合。阻遏蛋白的第 50 至第 80 氨基酸残基构成的铰链区将手臂与核心片断(第 81～360 氨基酸残基)连接起来。

既然头部片断从核心片断上伸展出来又具有结合 DNA 的能力，那末四聚体的功能又是什么？它们如何与 DNA 结合？一种模型认为，四个亚基的核心片断相互接触形成四聚体，而头部片断成对结合。在大约 115Å 长的哑铃状或圆柱状的四聚体分子的两端各有一对头部片断。四聚体与 *lacO* 结合时，二者的轴心以 15～20° 相交。四聚体的核心部分结合于 *lacO*的中心区域，而两对头部片断中可能只有一对与 DNA 结合。另一种模型与第一种模型大体相同，只是认为阻遏蛋白的轴心与 DNA 的轴心相互垂直。这样，也只能有一对头部片断能与 DNA 结合。

3. 阻遏蛋白与操作子的解离　诱导物进入细胞以后，是直接与操作子上的阻遏蛋白相结合还是与细胞中的自由四聚体相结合从而打破了结合状态四聚体与自由四聚体的平衡进而促进结合的四聚体与 DNA 的解离？前一种可以称为直接作用，后一种可以称为间接作用。大量证据表明，这种作用是直接作用而不是间接作用。首先，四聚体与操作子结合是相当稳定的，其半衰期大约为数十分钟，而诱导作用却要快得多。其次，虽然结合了诱导物(如 IPTG)的阻遏蛋白与*lacO*的亲和力很低，但人们通过用高浓度的阻遏蛋白-IPTG 复合物进行补偿试验，得到了与无诱导物存在时相同的甲基化试验结果(受保护的 DNA 核苷酸序列和促进甲基化的碱基位点，见前面所述)。用对操作子具有不同亲和力的突变阻遏蛋白(结合常数相差 7 个数量级)进行补偿试验也都得到相应的结果。这就说明，阻遏蛋白四聚体结合于操作子时，诱导物是能够与之结合的。诱导物结合位点位于核心片断上。结合了诱导物之后，核心片断发生了变构，这一作用通过铰链区再传导到头部片断，引起整个四聚体构象的改变，从而降低了与操作子的结合能力(下降约 1000 倍)，阻遏蛋白四聚体就离开了操作子。诱导物引起变构的假设模式示于图 8-6。

阻遏蛋白与其他许多位点特异性的 DNA 结合蛋白(如 RNA 聚合酶)一样，都具有以较低

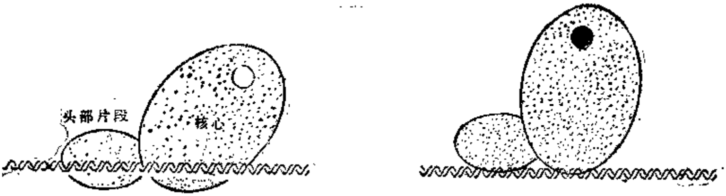



图 8-6　诱导物引起的变构作用削弱了四聚体与操作子的全部结合点。左图表示头部片断沿着DNA双螺旋与操作子特异性地结合；右图表示诱导物引起的变构改变了头部片断和核心片断之间的角度，从而使四聚体不再与操作子紧密结合

的亲和力与随机的DNA序列相结合的能力。因此，细胞内并不存在着游离的阻遏蛋白四聚体。阻遏蛋白四聚体与操作子的结合常数是与随机序列的结合常数的 $4\times10^6$ 倍。虽然随机序列对于阻遏蛋白的亲和力是如此之低，但是数量巨大是其与操作子竞争阻遏蛋白的重要因素。操作子结合阻遏蛋白的长度大约26bp，而大肠杆菌细胞内则有 $4.2\times10^6/26=1.6\times10^5$ 个非特异性结合位点，而特异的结合位点只有一个。因此，在没有诱导物存在的情况下，细胞内一个阻遏蛋白四聚体结合操作子的机率仅仅是结合非特异性序列的机率的25倍（$4\times10^6/1.6\times10^5=25$）。因此，任何使阻遏蛋白与操作子的特异性亲和力下降20～30倍的突变都能使 *lac* 酶系统的合成变为组成型。

在无诱导物存在时，细胞内有一个阻遏蛋白四聚体结合在操作子上，其余全部（或绝大部分）随机地结合在其他DNA序列上。加入诱导物以后，阻遏蛋白因变构作用使之与 *lacO* 的亲和力下降1000倍，但与非特异性序列的亲和力不变，因此，*locO* 上的阻遏蛋白四聚体迅速转移到随机的DNA序列上去。诱导物被代谢完毕后，一个阻遏蛋白四聚体又迅速回到操作子上。这一过程不大可能是沿着DNA滑动，而可能是由于DNA链的盘绕，而使某一部位上的四聚体直接进入操作子上的特异性结合位点。

4. 影响阻遏蛋白与操作子相互作用的突变体　由于每个阻遏蛋白亚基都具有两种结合位点，一种是与DNA结合的操作子位点，一种是与诱导物结合的诱导物位点。诱导物结合于诱导物位点以后就引起了阻遏蛋白分子的变构，失去与操作子特异结合的能力。因此野生型的 *lac* 操纵元是可诱导的。下面我们要讲到的突变体的表现型为显性或隐性，均是对于野生型而言的，即在部分二倍体中，总有一个野生型基因或调控位点。

（1）$i^-$突变体：突变位点分布在核心片断上，基因产物或者失去形成四聚体的能力，或者虽然能形成四聚体但不能与操作子结合。该突变基因对于野生型来说为隐性。在突变的单倍体中，Lac 酶系统的表达为组成型。

（2）$i^s$ 突变体：即超阻遏突变体，突变位点广泛分布于第62至第300个氨基酸残基上，特别是在第190至第300个残基范围内较为集中，而且呈现间隔为约26个残基的周期性分布。这类突变，或者是改变了诱导物结合位点，使阻遏蛋白不能结合诱导物，或者虽然能结合诱导物但不能产生变构效应（变构效应在传导过程中发生中断而不能到达操作子结合位点）。$i^s$ 突变对野生型为显性。在突变的单倍体中，不能诱导 Lac 酶系统的表达。

（3）$i^t$ 突变体：即紧密结合突变体，突变位点分布在前50个氨基酸的头部片断中，造成阻遏蛋白与操作子的紧密结合，因而也导致 Lac 酶系统的不可诱导性。

(4) $i^{-d}$ 突变体：突变位点分布在头部片断上。当这种突变基因产生的阻遏蛋白组成同种四聚体时，其效应与一般的 $i^-$ 突变体完全相同。而当 $i^{-d}$ 基因产物与野生型阻遏蛋白亚基组成异种四聚体时（理论上有 87.5% 的机率），则大约有 40% 的异种四聚体不能与 *lacO* 结合，也就是说对于野生型表现为显性。这样 Lac 酶系统就表现为组成型合成。其余 60% 的异种四聚体表现出野生型四聚体的行为。

(5) $O^c$ 突变体（见图 8-5）

现将 $i^-$、$i^s$、$i^{-d}$ 等突变体中突变位点的分布状况示于图 8-7。各种突变体对于野生型的关系（隐性或显性）以及对于 Lac 酶系统表达的影响列于表 8-2。

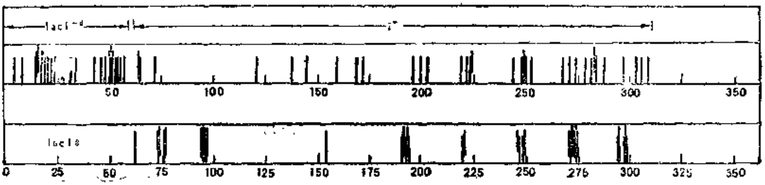



图 8-7　大肠杆菌LacI基因突变的分布

**表 8-2　影响阻遏蛋白与操作子相互作用的突变体及其效应**

| 突变体 | 突变点分布 | 顺式/反式 | 显性/隐性 | Lac 酶系统表达 |
|---|---|---|---|---|
| $i^+$ | | 反　式 | | 可诱导型 |
| $i^-$ | 核心片断 | 反　式 | 隐　性 | 组成型 |
| $i^s$ | 核心片断 | 反　式 | 显　性 | 不可诱导型 |
| $i^s$ | 头部片断 | 反　式 | 显　性 | 不可诱导型 |
| $i^{-d}$ | 头部片断 | 反式 同种 | 隐　性 | 组成型 |
| | | 反式 异种 40% | 显　性 | 组成型 |
| | | 反式 异种 60% | 中　性 | 可诱导型 |
| $O^c$ | 操作子 | 顺　式 | 显　性 | 组成型 |

## 四、乳糖操纵元的正调控

关于 cAMP-CAP 对于乳糖操纵元的正调控作用及其机制，我们在第六章中已经讲过。在遗传学方面，人们已经分离出两类使乳糖操纵元在无葡萄糖而有乳糖存在时也不能表达的突变体，一类位于编码 CAP 的 *cap*(*crp*) 基因上；另一类位于编码合成 cAMP 的腺苷酸环化酶(adenylate cyclase)基因 *cya* 上。细胞中 CAP 的合成是组成型的，因而也是相对稳定的。这样，具有正调控能力的 cAMP-CAP 复合物主要取决于 cAMP 的含量。腺苷酸环化酶位于细胞膜上。该酶的活性与细胞膜的组织结构和生理状态，特别是与负责葡萄糖运输的磷酸烯醇式丙酮酸磷酸基转移酶的活性有关。因此，cAMP-CAP 不仅调控乳糖、半乳糖、阿拉伯糖等糖类代谢有关的酶，而且三羧酸循环和呼吸链酶系统中的大多数酶，分解各种碳源底物的酶，降解某些氨基酸的酶，乙醛酸途径的酶等，均表现为对葡萄糖效应敏感。

CAP 为二聚体蛋白质，亚基分子量为 22.5 KDa，在启动子上的结合位点一般都包含

TGTGA，在乳糖操纵元中为对称序列，但大多数为非对称序列。这可能反映了 cAMP-CAP 系统对不同的操纵元有不同程度的控制。不同的操纵元中 CAP 位点与转录起始点的距离相差很大：在乳糖操纵元中位于 −72～−52；半乳糖操纵元中 −76～−47；阿拉伯糖操纵元中为 −107～−78。人们发现，缺乏 −35 序列的操纵元对 cAMP-CAP 系统的依赖性更大。依赖 cAMP-CAP 的操纵元都没有很好的 −35 序列，有的甚至还没有好的 −10 序列。因此 cAMP-CAP 与其结合位点的相互作用，RNA 聚合酶与 −35 序列的相互作用，二者是相关的，都具有促进转录起始的作用。促进转录的可能机制是促进 DNA 的解链，使开放性启动子复合物能够形成。不但减少 −10 序列中 G·C 对的突变能够降低转录起始对 cAMP-CAP 系统的依赖性(见第六章)，而且 DNA 拓扑异构酶 I (能减少负超螺旋)的突变也能降低对 cAMP-CAP 系统的依赖。当然这也并不排斥 cAMP-CAP 复合物以某种方式促进 RNA 聚合酶与启动子结合的可能性(见第十章)。

## 第四节　操纵元的其他调控形式[5,6,8~13]

### 一、具有双启动子的半乳糖操纵元

半乳糖也是被细菌用作碳源的一个糖类。细胞代谢半乳糖，需要三种酶，即半乳糖激酶，半乳糖转移酶，半乳糖表异构酶。在细胞内的代谢步骤如下：

$$\text{半乳糖} + \text{ATP} \xrightarrow{\text{半乳糖激酶}} \text{半乳糖 1-磷酸} + \text{ADP} + \text{H}^+$$

$$\text{半乳糖 1-磷酸} + \text{UDPGlc} \xrightarrow{\text{半乳糖转移酶}} \text{UDPGal} + \text{葡萄糖 1-磷酸}$$

$$\text{UDPGal} \xrightarrow{\text{半乳糖表异构酶}} \text{UDPGlc}$$ 以上三个反应式的总反应是：

$$\text{半乳糖} + \text{ATP} \longrightarrow \text{葡萄糖 1-磷酸} + \text{ADP} + \text{H}^+$$

上述反应式中 UDPGlc 为尿苷二磷酸葡萄糖，UDPGal 为尿苷二磷酸半乳糖。上述三个酶的基因即 galK，galT，galE 组成半乳糖操纵元的主体部分，而编码阻遏蛋白的基因 galR 则相距很远，这是 *gal* 操纵元与 *lac* 操纵元显著的差异 (图 8-8)。实际上并没有理由期待调节基因一定要靠近操作子，因为调节基因的产物是可扩散的。因此实际上大肠杆菌中调节基因有的离操作子很近，有的则很远。半乳糖操纵元的阻遏蛋白作用方式与 *lac* 阻遏蛋白一样，*gal* $R^-$ 和 *gal* $O^c$ 都造成 Gal 酶系统的组成型表达。*gal* 操纵元负控制系统的诱导物是半乳糖而不是其代谢产物。许多分解代谢酶系统的操纵元都是如此。

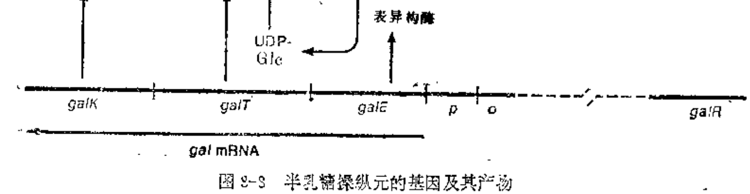



图 8-8　半乳糖操纵元的基因及其产物

细菌对半乳糖的利用要比葡萄糖差得多。人们可能会猜想，当葡萄糖存在时，半乳糖操纵元是不会被诱导的。然而实际上当葡萄糖存在时，*gal* 操纵元仍可被诱导(虽然水平低得多)。通过对启动子突变的分析，发现 *gal* 操纵元有两个相互重叠的启动子。*galP1* 是依赖于cAMP-CAP 的，转录起点为 S1；*galP2* 是不依赖 cAMP-CAP 的，转录起点为 S2。两个启动子各有自己的 Pribnow 框，两者相差 5 bp(图 8-9)。

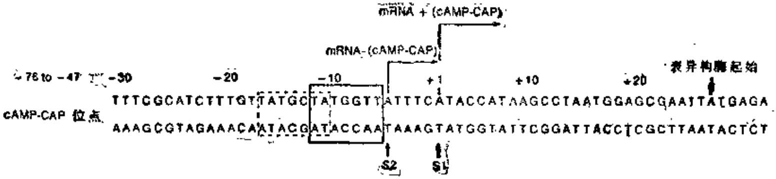
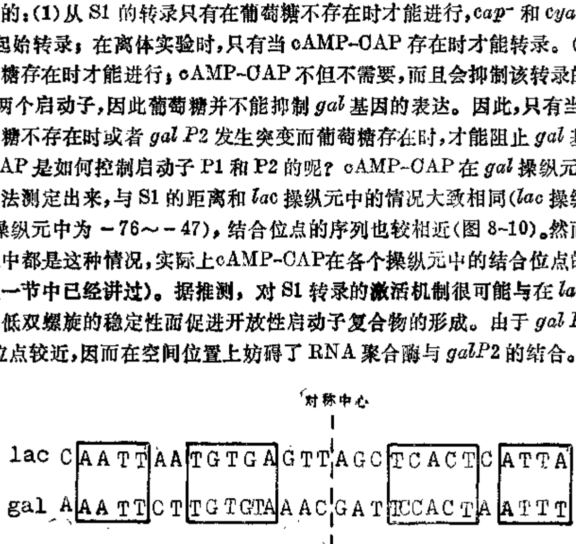



图 8-9　*gal* 操纵元的操作子和启动子碱基序列。S1 为依赖 cAMP-CAP 的 mRNA 的起点，S2 为不依赖于 cAMP-CAP的 mRNA 的起点。

人们通过测定在不同的条件下得到的 *gal* mRNA5′ 末端的序列，证明了 *gal P1* 和 *galP2* 的功能是不同的：(1) 从 S1 的转录只有在葡萄糖不存在时才能进行，$cap^-$ 和 $cya^-$ 两种突变体都不能从 S1 起始转录；在离体实验时，只有当 cAMP-CAP 存在时才能转录。(2) 从 S2 的转录只有当葡萄糖存在时才能进行；cAMP-CAP 不但不需要，而且会抑制该转录的起始。由于 *gal* 操纵元有两个启动子，因此葡萄糖并不能抑制 *gal* 基因的表达。因此，只有当 *gal P1* 发生突变而且葡萄糖不存在时或者 *gal P2* 发生突变而葡萄糖存在时，才能阻止 *gal* 基因的表达。

cAMP-CAP 是如何控制启动子 P1 和 P2 的呢？cAMP-CAP 在 *gal* 操纵元上的结合位点已通过脚印法测定出来，与 S1 的距离和 *lac* 操纵元中的情况大致相同(*lac* 操纵元中为 −72～−52，*gal* 操纵元中为 −76～−47)，结合位点的序列也较相近(图 8-10)。然而这并不是说在每个操纵元中都是这种情况，实际上cAMP-CAP在各个操纵元中的结合位点的位置和序列差异很大（上一节中已经讲过)。据推测，对 S1 转录的激活机制很可能与在 *lac* 操纵元中一样，是通过降低双螺旋的稳定性而促进开放性启动子复合物的形成。由于 *gal P2* 离 cAMP-CAP 的结合位点较近，因而在空间位置上妨碍了 RNA 聚合酶与 *galP2* 的结合。



*galO* 位点，称为 $galO_I$（原来 -60左右的 *galO* 称为 $galO_E$）。$galO_E$ 与 $galO_I$ 相距90bp以上。$galO_I$ 的作用机制也不清楚。$galO_I$ 的存在能够加强 $galO_E$ 对于转录的阻遏作用。这可能是通过两个操作子之间的 DNA 形成噜噗结构，两个阻遏蛋白分子相互作用而强化了阻遏。当 gal 阻遏蛋白存在时(即缺乏半乳糖时)，从P2常可以起始转录，但产生 20bp 左右的 mRNA便停止转录，这可能与 $galO_I$ 与 $galO_E$ 的相互作用有关。

为什么 *gal* 操纵元需要两个启动子呢？这与半乳糖在分解代谢和合成代谢中的双重作用有关。一方面半乳糖可以作为细菌的碳源，而另一方面，UDPGal是合成细菌细胞壁的一个重要前体。无外源半乳糖时，细菌必须将自己的 UDPGlc 在半乳糖表异构酶的作用下转变为 UDPGal，而这个酶正是 *gal*E 所编码的。因而细菌必须以低水平不断地合成 GalE 才能维持其生长。这种本底组成型的转录是由 *galP*2 所启动的，它不因葡萄糖的存在而削弱。让我们设想一下，如果只有 *galP*1存在，当有葡萄糖时就不能合成 GalE 了；而如果只有 *galP*2 存在，在没有葡萄糖的情况下，细菌也不能利用半乳糖，因为既不能高水平地合成 Gal 酶系统(因缺少 *galP*1)，又因为 cAMP-CAP 抑制了 *gal*P2，低水平合成 Gal 酶系统也是不可能的。因此，有了双启动子，既能满足经常的低水平的需要，又能满足临时的大量需要。

## 二、阿拉伯糖操纵元：具有双重功能(正控制和负控制)的调节蛋白质

阿拉伯糖是一个可以用作碳源的五碳糖。参与阿拉伯糖代谢的酶基因分为三个操纵元，一个是 *araBAD*，一个是 *araE*，另一个是 *araFG*，由一个共同的 *araC* 基因产物进行调控。这种由一个调节基因控制几个操纵元的系统，称为调控元（regulon）。*araE* 和 *araFG* 编码的膜结合蛋白，与阿拉伯糖结合和转运有关，下文就不再讨论。而 *araBAD* 分别编码 *L*-核酮糖激酶，*L*-阿拉伯糖异构酶和 *L*-核酮糖-5-磷酸-4-表异构酶。*araBAD* 的基因组成和它们所编码的酶所催化的反应示于图 8-11。

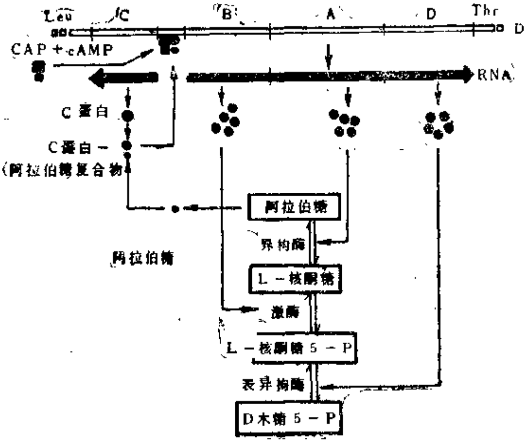



图 8-11　大肠杆菌 *araBAD* 操纵元的基因产物所催化的反应及其基因组成

*araC* 与 *araBAD* 的转录方向相反，因此它们都有各自的启动子($araP^{C}$ 和 $araP^{BAD}$)。

*araC* 基因产物对 *araBAD* 具有正调控作用。其遗传学证据如下：

(1) 所有的点突变和缺失突变 ($araC^{-}$) 都造成 *araBAD* 的不可诱导性。但是不是 *araC* 产物为 *araBAD* 的阻遏蛋白，而 $AraC^{-}$ 失去了结合诱导物的能力？如果确是如此，则 $araC^{-}$ 对 $araC^{+}$ 应该是显性的。然而回答是否定的，下一个实验则说明这个问题。

(2) $F'araC^{+}/araC^{-}$ 这样的部分二倍体则表现出完全的可诱导性。这就说明 $araC^{-}$ 对于 $araC^{+}$ 为隐性。但 *araC* 是否可能是 *araBAD* 启动子的一部分呢？下一个实验也否定了这种可能性。

(3) $F'araC^{+}araA^{-}/araC^{-}A^{+}$ 则表现出完全可诱导性。这就说明 *araC* 不是顺式作用成分，而是编码可扩散的反式作用因子。下一个实验则证明了这个因子是一种蛋白质。

(4) 在 $araC^{-}$ 突变中，有些为 Amber 突变，因此该基因产物为蛋白质。现已分离出该蛋白质。AraC 是 *araBAD* mRNA 合成所不可缺少的正调控因子。

原来AraC蛋白质具有两种形式，一种是在没有阿拉伯糖存在时，AraC表现为对 *araBAD* 和 *araC* 的阻遏作用。这种形式的 AraC 称为 $C^{rep}$。当有阿拉伯糖存在时，AraC 便与阿拉伯糖结合，表现为对 *araBAD* 的诱导作用。这种结合了阿拉伯糖的 AraC 称为 $C^{ind}$。然而$C^{ind}$也可以对 *araC* 产生阻遏作用。

最初，人们只发现了一个 *araO6* 位点，位于 −110～−140 之间，此处正是 *araC* 基因的聚合酶结合位点，因此阻遏蛋白 AraC 在这个操作子上的结合有效地阻遏了 *araC* 基因的转录。而这种结合对于 *araBAD* 转录只表现出轻微的阻遏作用。最近，Schleif 等人发现了另一个 *araO* 位点，位于 −280 附近(−265～−294)，即在基因 *araC* 的前导序列中(这段序列是转录的，但不翻译)。这个位点就称为 $araO_2$。它对于 *araBAD* 的有效阻遏是不可缺少的。在 $araO_2$ 与 *araBAD* 的启动子之间插入(或缺失)−16、−8、0、5、11、15、24、31 个 bp，结果只有 0、11、31bp 三种插入突变基本不影响对 *araBAD* 启动子的阻遏，其余均不同程度地降低了阻遏效应。不难看出，插入 11bp 和 31bp 均为整数螺旋圈数，不改变 $araO_2$ 上的阻遏蛋白及与其相互作用的蛋白质在 DNA 双螺旋侧面上的位置关系。这就表明 AraC 与 $araO_2$ 结合以后，通过形成嗜喋结构，再与 *araBAD* 启动子上的某种蛋白质 (CAP, AraC, RNA 聚合酶) 相互作用，实现了有效的阻遏。据认为，一个 AraC 二聚体同时结合于 $araO_2$ 和 *araI* 而实现了对 *araBAD* 的阻遏。当没有 cAMP-CAP 时，即使有阿拉伯糖存在，也不能诱导 *araBAD* 的转录。

cAMP-CAP 也是 *araBAD* 和 *araC* 表达所必需的。因此 *araBAD* 的表达需要两个正调控因子，一个是 $C^{ind}$，另一个就是 cAMP-CAP。当有葡萄糖存在时，即没有 cAMP-CAP，*araC* 也能够以很低的本底水平表达，但很快就会被 $C^{rep}$ 所阻遏。这种作用就叫自身调控 (autogenous regulation)。$C^{rep}$ (或 $C^{ind}$) 对 *araC* 的阻遏作用是由于 $C^{rep}$ ($C^{ind}$) 作用于 $araO^{C}$(−144～−106)，而 $araO^{C}$ 与 $araP^{C}$ 相互重叠，因此 $C^{rep}$(或 $C^{ind}$)的结合妨碍了 RNA 聚合酶的结合。而 $C^{ind}$ 对于 *araBAD* 的激活是由于 $C^{ind}$ 结合于 *araI*(−78～−40)。这一位置下游方向与 RNA 聚合酶结合位点紧密相连，而上游方向与 cAMP-CAP 位点 (−107～−78)紧密相连。当 cAMP-CAP 结合上来以后，可同时激活 *araBAD* 和 *araC*。当 *araC*表达过多时，则由 $C^{ind}$ 结合于 $araO_c$，从而阻遏了 *araC* 的表达。若有足够量的 *araBAD* 酶系统产生出来，则 $C^{ind}$ 减少或消失，这样 *araBAD* 则被关闭。图 8-12 就是 *ara* 操纵元调控机制的

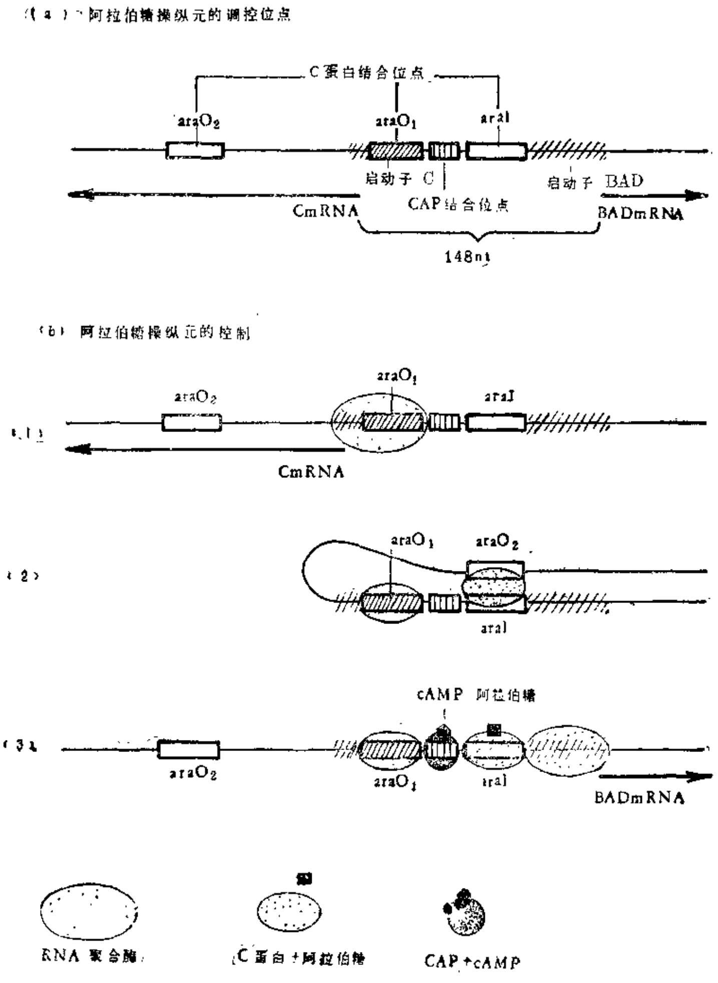



图 8-12 *E.coli* 阿拉伯糖操纵元的调控

(a) 阿拉伯糖操纵元的调控位点。(b) 阿拉伯糖操纵元的控制：(1) 在没有C蛋白时，启动子BAD不被激活；(2) 在有C蛋白存在而cAMP含量低时，无论有、无阿拉伯糖存在，启动子BAD受到阻遏，启动子C也受到阻遏；(3) 在有C蛋白存在cAMP含量高且有阿拉伯糖存在时，启动子BAD才被激活，启动子C仍然被阻遏。

模型。

*araBAD* 的表达必须在 $C^{ind}$ 和 cAMP-CAP 同时存在时才能进行。突变体的研究表明，−35 序列的突变（*araI*$^{C}$）可使 *araBAD* 在无 $C^{ind}$ 时表达；而 −10 序列的突变（*araX*$^{C}$）则可使这些基因在无 cAMP-CAP 时表达。−35 序列和 −10 序列都位于 RNA 聚合酶结合区域（*araP*$^{BAD}$），说明 $C^{ind}$ 和 cAMP-CAP 对于 *araBAD* 转录的激活作用在于促进 RNA 聚合酶与 *araBAD* 的结合并形成开放性启动子复合物。最近的研究表明，AraC 是以二聚体形式与 ara*I* 作用的，它与以后要讲的 λ 阻遏蛋白不同，这种二聚体能作用于相邻的三个大沟。它与 RNA 聚合酶的直接接触的机制还不清楚，cAMP-CAP 的作用机制可能与它在 *lac* 操纵元中的作用机制相似。

### 三、色氨酸操纵元：可阻遏系统

以上所讲的 *lac*，*gal*，*ara* 三种操纵元都是可诱导的系统，它们负责某一营养基质的分解。而它们的诱导物就是需要分解的底物或其变构形式。细菌中还有负责某些物质合成的操纵元。比如负责氨基酸合成的操纵元。在没有外源氨基酸时，这类操纵元表达，使细胞内有足够的氨基酸以进行蛋白质合成。如果有外源氨基酸存在，则细菌就不必自己合成，这类操纵元就受到阻遏。这类可以被最终合成产物所阻遏的操纵元就叫做可阻遏操纵元（repressible operons）。色氨酸操纵元就是一个典型的可阻遏操纵元。

*trp* 操纵元的基因组成示于图 8-13，它含有五个基因。*trpE* 和 *trpD* 分别编码邻氨基苯甲酸合成酶（anthranilate synthetase）两个亚基，各为 60KDa；*trpC* 编码吲哚甘油磷酸合成酶（indoleglycerol-phosphate synthetase），分子量为 45 KDa；*trpB* 和 *trpA* 分别编码色氨酸合成酶的 β 亚基（50KDa）和 α 亚基（29KDa）。在 *trpO* 与 *trpE* 之间有一段 162bp 的前导序列可以转录到 mRNA 中。在 *trpA* 基因下游方向 36bp 处于一个不依赖于 ρ 因子的终止子 t，在 t 的下游 250bp 处还有一个依赖于 ρ 因子的终止子 t′。与 *trpO* 结合的阻遏蛋白由相距甚远的 *trpR* 编码。

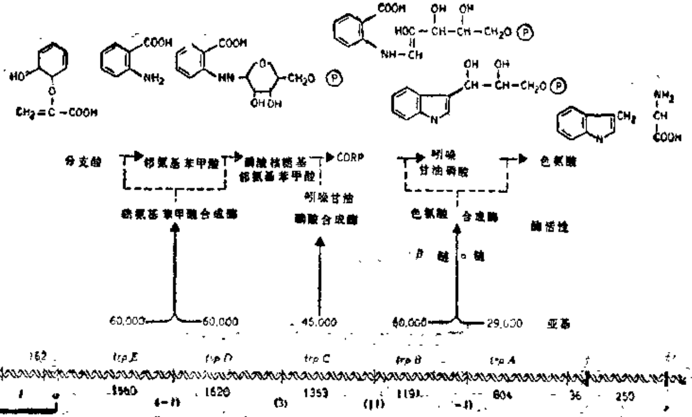



图 8-13　大肠杆菌色氨酸操纵元的基因组成及其基因产物所催化的反应

P：启动子；O：操作子；L：前导序列；a：衰减子；t，t′：终止子。括号内数字表示基因间距离（碱基对）。

色氨酸合成酶系统的活性和酶的合成都受色氨酸的控制。一方面，色氨酸对已有的酶起反馈抑制(feedback inhibition)，另一方面，色氨酸作为辅阻遏物激活无活性的无辅基阻遏蛋白(*trpR* 基因产物)，从而结合到操作子 *trpO* 上，阻止 *trp* 操纵元的转录(图 8-14)。

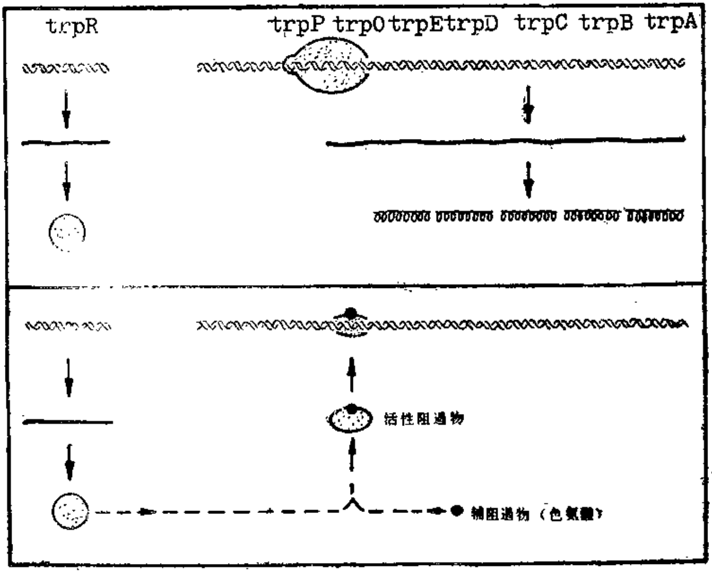



图 8-14　色氨酸操纵元的可阻遏调控系统

*trpR* 编码 12.5KDa 的阻遏蛋白亚基，在细胞内以四聚体的形式起作用。每个细胞内约有 20 个这样的四聚体。单独的四聚体不能和 *trpO* 结合；只有先结合了色氨酸分子后改变了四聚体的分子构象才能与 *trpO* 结合。

*trpP* 位于 −40 到 +18 这个范围，而 *trpO* 整个位于 *trpP* 内(−21～+1)并有比较完美的 20bp 的反向重复序列。因此，活性阻遏物与 *trpO* 的结合则完全排斥了 RNA 聚合酶的结合。*trpP* 有正常的 −10 序列和 −35 序列，而 −10 序列完全在 *trpO* 之内。*trpP* 和 *trpO* 的碱基序列示于图 8-15。

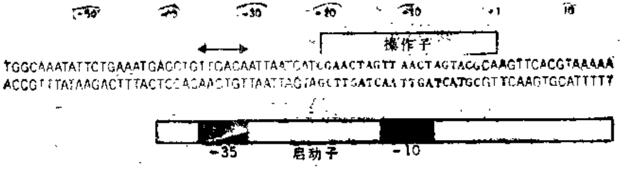



图 8-15　trp 操纵元的操作子完全位于启动子内

## 第五节　基因转录的时序调控[14~38]

原核生物在生长发育的各个阶段，如细胞分裂，芽孢形成，噬菌体的复制和噬菌体颗粒的装配等，基因的表达是按照一定时间顺序而展现的。这种调控机制统称为时序调控。由于基因表达的时序调控，使得生物能够有效地利用能源，避免浪费，同时也避免了某些基因在不适当时机表达可能带来的危害。

基因表达的时序调控，大多是通过一种或几种蛋白质因子与RNA聚合酶相互作用而实现的。这些蛋白质因子，有的替代原来的 $\sigma$ 亚基以协助核心酶识别特定的启动子。如枯草杆菌噬菌体SPO1的早、中、晚三个阶段基因表达过程中的 $\sigma$ 亚基的替换和枯草杆菌芽孢形成过程中的 $\sigma$ 亚基替换以及大肠杆菌热震惊基因的表达。有的则修饰RNA聚合酶的核心酶并同时更换 $\sigma$ 亚基，如T4噬菌体早、晚期基因的表达。如果缺乏 $\sigma$ 亚基替换和对核心酶进行修饰手段，有的干脆使原有的RNA聚合酶失活而代之以新的RNA聚合酶，如T7噬菌体三类基因转录的调控（利用这种机制的也只可能是噬菌体）。有的蛋白质因子影响到RNA聚合酶与转录终止子的作用，从而控制不同阶段基因的表达，如 $\lambda$ 噬菌体的抗终止因子（见第六章）。而 $\lambda$ 噬菌体的CⅡ蛋白和CⅢ蛋白对于从 $P_E$ 转录 $cI$ 基因和从 $P_I$ 转录整合酶基因的促进作用又是以一种正调控因子的身份协助RNA聚合酶对于这两个启动子的识别。

### 一、枯草杆菌中 $\sigma$ 的亚基更迭

**1．枯草杆菌噬菌体SPO1早、中、晚基因表达的转换**　SPO1为一种烈性噬菌体，其生活史可以分为早、中、晚三个时期。在侵染后，早期基因立即被转录。执行这个任务的酶是寄主的RNA聚合酶全酶 $\alpha_2\beta\beta'\sigma55$。与此相应的是早期基因的启动子的－10序列和－35序列与寄主的一致序列很相似。在侵染4～5分钟后，早期基因停止表达而开始转录中期基因。而从侵染后的8分钟至12分钟，则中期基因转录又被晚期基因的转录所代替。

SPO1是怎样实现从早期基因的转录到中期基因的转录进而到晚期基因转录的两次转变过程呢？SPO1中、晚期基因的表达需要三种SPO1自身的调节基因的表达，这三种调节基因是基因28、33、34。基因28为早期基因。寄主的RNA聚合酶全酶转录早期基因后就产生调节蛋白gp28，其分子量为26KDa。gp28取代了 $\sigma55$ 并与核心酶结合，从而改变RNA聚合酶对于启动子的识别特性，即不再识别早期基因启动子而只识别中期基因的启动子。中期基因33和34的突变导致晚期基因不能表达，因此这两个基因产物gp33和gp34又是晚期基因表达的调节蛋白。gp33和gp34的分子量分别为13KDa和24KDa，它们可以取代gp28（以及 $\sigma$ 55）并使RNA聚合酶只能识别晚期基因。这些调节蛋白质之所以能一个取代另一个，可能是由它们与核心酶的亲和力所决定的。这种连续地更换 $\sigma$ 亚基的途径使SPO1的早、中、晚三个阶段的基因能够有条不紊地表达（图8-16）。这种 $\sigma$ 亚基的更换可能是彻底的，因而这种变化是不可逆的。

**2．枯草杆菌孢子生成过程中 $\sigma$ 亚基的替换**　大肠杆菌在通常情况只有一种 $\sigma$ 亚基，而在枯草杆菌中更换 $\sigma$ 亚基则是常见的现象。除了前面所说的噬菌体SPO1更换 $\sigma$ 亚基之外，枯草杆菌自己在芽孢形成的过程中也采用更换 $\sigma$ 亚基的战略。图8-17表示在孢子形成过程中主要的 $\sigma$ 亚基更换事件。$\sigma^{55}$，$\sigma^{37}$，$\sigma^{29}$ 分别为分子量55KDa、37KDa、29KDa的三种 $\sigma$ 亚

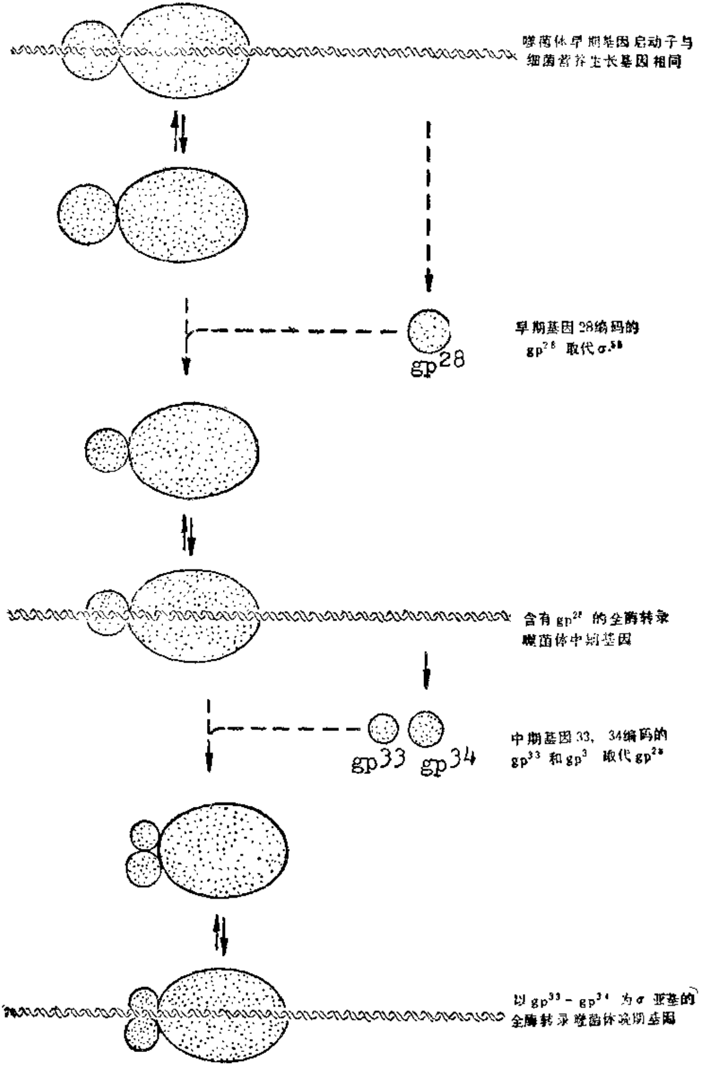



图 8-16 噬菌体 SPO1 通过先后两次更换 $\sigma$ 亚基以实现早、中、晚三个时期基因表达的调控

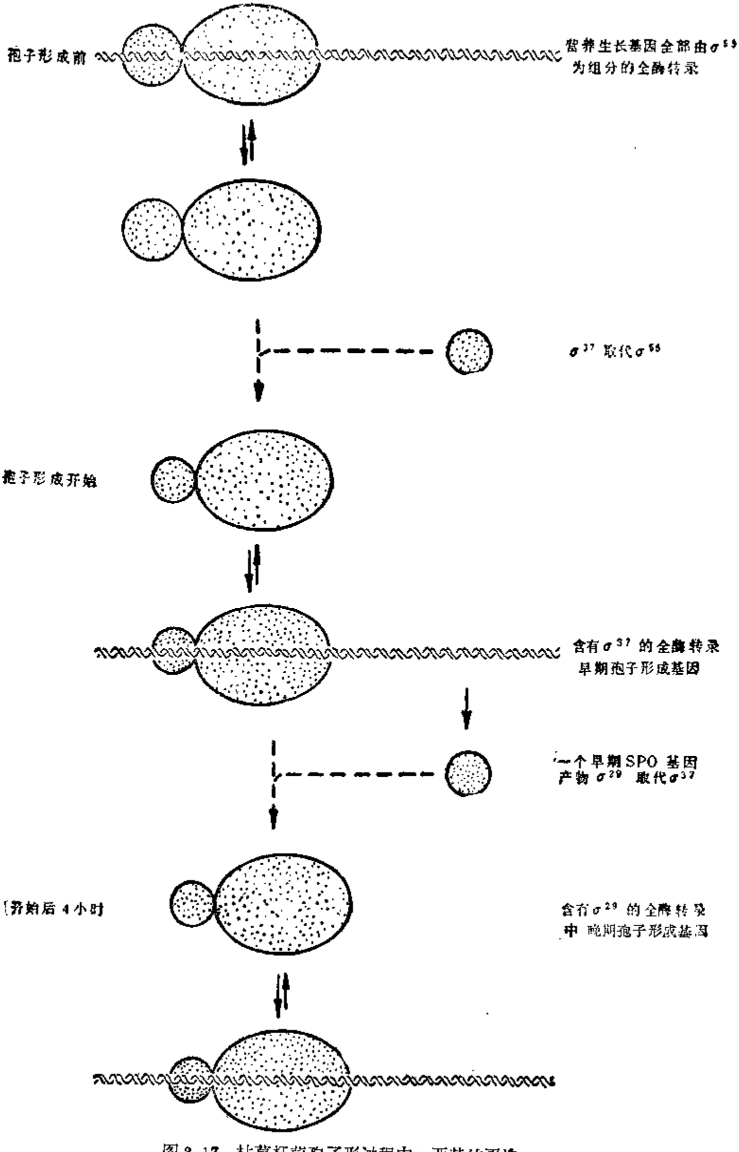



图 3-17 枯草杆菌孢子形过程中σ亚基的更迭

基，含有 $\sigma^{55}$ 的 RNA 聚合酶只能识别营养基因 (vegetative genes)的启动子，含有$\sigma^{37}$ 的聚合酶能识别早期孢子形成基因的启动子，而含有 $\sigma^{29}$ 的聚合酶则负责中、晚期孢子形成基因的转录。需要指出，枯草杆菌本身 $\sigma$ 亚基的代换并不是彻底的代换，例如含有 $\sigma^{37}$ 的 RNA 聚合酶只占 RNA 聚合酶总分子数的 10%。

以上三种 $\sigma$ 亚基中，只有 $\sigma^{55}$ 存在于营养生长阶段的细胞中，负责与营养生长有关基因的转录和孢子形成过程最初所需要的基因 (*spoO* 基因)。在 *spoO* 基因中，有一个基因的产物就是 $\sigma^{28}$。在营养生长阶段的细胞中，含有 $\sigma 28$ 的 RNA 聚合酶只占 RNA 聚合酶总分子数的很小很小的比例；而且，一旦孢子生成过程开始，$\sigma^{28}$ 即失去活性。$\sigma^{28}$ 有可能是信号系统的一个成员，负责转录那些能够探测营养耗竭和起始孢子形成反应的基因。另外在营养生长阶段的细胞中，还发现了 $\sigma^{32}$，含有 $\sigma^{32}$ 的 RNA 聚合酶比含有 $\sigma^{28}$ 的要多，但也只占总数的 1% 以下。同样，$\sigma^{32}$ 也负责识别某些基因的启动子，它在孢子形成过程的早期才有活性；现在还不能确定枯草杆菌的芽孢形成过程中到底有多少种 $\sigma$ 亚基；就目前已发现的亚基来说，更迭的次序大概是 $\sigma^{55}$、$\sigma^{28}$、$\sigma^{32}$、$\sigma^{37}$、$\sigma^{29}$。

这些不同的 $\sigma$ 亚基，包括枯草杆菌本身的和噬菌体 SPO1 的，都能够识别特定的启动子。$\sigma^{55}$ 所识别的启动子，具有与 *E.coli* 大体相同的 －10 序列和 －35 序，而其他 $\sigma$ 亚基所识别的启动子都不具备 $\sigma^{55}$ 所识别的启动子结构(－10 序列和 －35 序列)。而这些 $\sigma$ 亚基都有各自的启动子结构特征(表 8-3)。

以上 7 种 $\sigma$ 亚基能够识别 7 种不同类型的启动子。看来不像是 $\sigma$ 亚基影响了 RNA聚合酶全酶的构象，而可能由这些 $\sigma$ 亚基直接识别启动子的特征。目前还不能排除核心酶亚基有被修饰的可能性，就像 T4 噬菌体在侵染大肠杆菌后所使用的战略。

**表 8-3　枯草杆菌中每个 $\sigma$ 因子都能识别具有特征性的一致序列的一组启动子**

| $\sigma$ 因 子 | 用　　途 | －35 序 列 | －10 序 列 |
|---|---|---|---|
| $\sigma^{55}$ | 营养生长 | TTGACA | TATAAT |
| $\sigma^{28}$ | 孢子生成的信号系统 | CTNAAA | CCGATAT |
| $\sigma^{32}$ | 孢子生成 | AAATC | TANTGTTNTA |
| $\rho^{37}$ | 孢子生成 | AGGNTTT | GGNATTGNT |
| $\sigma^{29}$ | 孢子生成等 | TTNAAA | CATATT |
| $gp^{28}$ | SPO1 中期基因表达 | AGGAGA | TTTNTTT |
| $gp^{33-34}$ | SPO1 晚期基因表达 | CGTTAGA | GATATT |

注：本表中只有 $\sigma^{55}$ $gp^{28}$ $gp^{33-34}$ 所识别的一致序列是从相当多的启动子统计而来，其余的 $\sigma$ 因子所识别的序列所依据的资料不多。

## 二、大肠杆菌热震惊基因的表达

大肠杆菌在通常的生长条件下，是没有 $\sigma$ 亚基替换现象的。然而，如果让大肠杆菌生长在较高的温度下 (如 42°～50℃)，某些基因则迅速地表达。这种现象称为热震惊反应 (heat shock response)；在热震惊反应中大量表达的基因称为热震惊基因。不单是热能够引起热震惊反应，而且其他许多因素如高盐环境，重金属离子，极度干燥等均能引起这种反应。从广义来说，这些反应应该统称为应急反应 (stress response)。其中，人们对热震惊反应研究得最多。从大肠杆菌直到人类都具有热震惊基因。有些热震惊基因产物的氨基酸序列在不同种属中高度相关。甚至是核苷酸序列也有相关性，如分子量 66KDa 的 DnaK 的基因有近乎一半的核苷酸序列与果蝇的热震惊蛋白 Hsp70 的基因相同。

为什么在高温条件下热震惊基因能够迅速地表达?关于这一转变机制的关键性发现是:在大肠杆菌中存在着一种含量很少的分子量为32KDa的蛋白质。这种蛋白质是一种σ亚基,称为$\sigma^{32}$。$\sigma^{32}$由基因*rpoH*(过去称为*htpR*)所编码。由$\sigma^{32}$参与构成的RNA聚合酶全酶能够识别热震惊基因的启动子;由$\sigma^{32}$识别的这类基因的启动子与"标准"σ因子($\sigma^{70}$)所识别的启动子不同。在其－35序列的前面还有几个保守的碱基(TNNCNCNC),－10序列前面有保守的CCCC(见表8-4)。

表8-4 若干热震惊基因的启动子及其一致序列

| 启动子 | －35 区域 | 间隔区 | －10 区域 +1 |
|---|---|---|---|
| *groE* | TTTCCCCCTTGAA | GGGGCGAAGCCAT | CCCCATTTCTCTGGTCAC |
| *dnak P1* | TCTCCCCCTTGAT | GACGTGGTTTACGA | CCCCATTTAGTAG TCAA |
| *dnak P2* | TTGGGCAGTTGAA | ACCAGACGTTTCG | CCCCTATTACAGACTCAC |
| C62.5基因 | GCTCTCGCTTGAA | ATTATTCTCCCTTGT | CCCCATCTCTCCCACATC |
| $\sigma^{70}$热震惊启动子 | TGCCACCCTTGAA | AAACTGTCGATGTGG | GACGATATAGCAG ATAA |
| $\sigma^{32}$识别的一致序列 | T－－C－C－CTTGAA | 13～15bp | CCCCAT－T |
| $\sigma^{70}$识别的一致序列 | TTGACA | 15～20bp | TATAAT |

尽管有了$\sigma^{32}$可以识别热震惊基因的启动子,但这类基因表达的调节机制仍然不太清楚,可能也需要转录因子的活化和与其结合位点的结合。当细菌适应了较高生长温度时(如大肠杆菌在42℃),大多数热震惊基因便停止表达(如大肠杆菌在42℃时经过20分钟),又开始表达"常规基因。这中间的奥秘,仍然毫无所知。人们已知在热震惊反应时细胞内蛋白质降解速度急剧增加,其中可能包括$\sigma^{70}$的降解。在正常情况下,$\sigma^{70}$基因作为σ操纵元(包括S21基因,DNA引发酶基因和$\sigma^{70}$基因)的一部分而转录在多基因的mRNA中。然而,$\sigma^{70}$基因本身还有一个热震惊启动子,位于其前一个基因(DNA引发酶基因)内(图8-18)。尽管热震惊反应并不需要$\sigma^{70}$,但$\sigma^{70}$基因在热震惊反应中仍大量表达。这一情况可能与细菌适应较高温度后开始表达"常规"基因有关。我们曾在第四章的SOS修复机制中讲到过,在SOS反应中*lexA*基因大量表达。这两种情况有某种相似之处:细胞在不利条件下仍然要为恢复正常秩序作好准备。

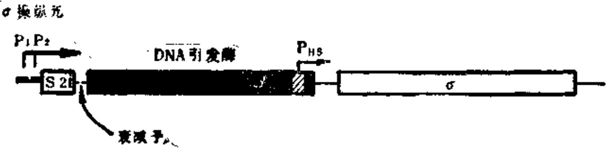



图8-18 σ操纵元的启动子(P1,P2)和$\sigma^{70}$基因的热震惊启动子($P_{HS}$)

## 三、T4噬菌体生长中RNA聚合酶亚基的修饰和σ亚基的替换

T4噬菌体为*E. coli*的烈性噬菌体,其DNA有166 000 bp,T4噬菌体的生长与SPO1相似,不同的生长阶段需要表达不同的基因。早期基因编码与DNA复制、转录调控、DNA重组有关的酶。晚期基因编码噬菌体衣壳蛋白以及噬菌体颗粒装配和细胞裂解所需的酶。在早期基因中,又可以分为几批,每批里有彼此相似的启动子结构。在T4侵染寄主后,寄主RNA聚合酶只能识别第一批早期基因启动子。在T4噬菌体头部还包含几种小的蛋白质分子,称为内部蛋白质。其中一种就是ADP核糖基转移酶(ADPRT)。在噬菌体侵染寄主时,ADPRT随着

DNA 一同进入寄主细胞。ADPRT 可以将一个或两个 ADP 核糖基连接到 RNA 聚合酶 $\alpha$ 亚基的精氨酸胍基上，经过修饰的 $\alpha$ 亚基使核心酶对于寄主 $\sigma$ 亚基的亲和力降低，而与 T4 基因所编码的蛋白质 Mot 的亲和力提高。Mot 取代了 $\sigma$ 亚基，降低了 RNA 聚合酶对寄主基因和 T4 本身第一批早期基因的识别能力，同时又能转录下一批基因。此后，RNA 聚合酶还要经历一系列的变化。到转录晚期基因时，RNA 聚合酶已被高度修饰：两个 $\alpha$ 亚基分别结合了一个 ADP-核糖基，另外还有四个蛋白质因子结合于酶分子。

高度修饰的 RNA 聚合酶本身仍不能转录晚期基因，而只有在 DNA 复制时才能转录这些基因。晚期基因所编码的蛋白质，只有当 DNA 复制到一定时候才需要，而且二者需要保持一定的数量关系。而 DNA 复制对这些蛋白质合成的调控容易满足这种定量关系。DNA复制对于晚期基因表达的调控机制还不清楚，可能因为复制改变了晚期基因的构象，使之易于起始转录。

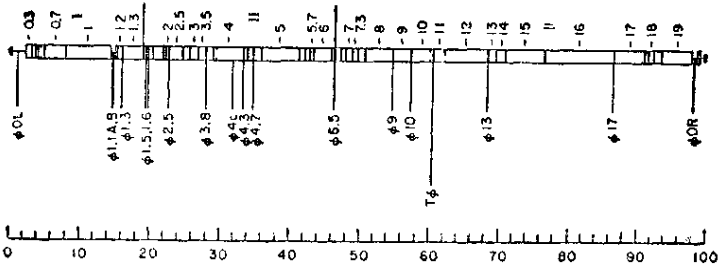



图 8-19　T7 噬菌体的遗传图

第 I 组(早期)基因：0.3，寄主限制酶的抑制剂；0.7，蛋白质激酶；1. T7RNA 聚合酶；1.2，复制蛋白质；1.3，DNA 连接酶。第 II 组基因（主要与 DNA 复制有关）：2，寄主 RNA 聚合酶抑制剂；2.5，DNA 结合蛋白；3，内切酶；3.5，溶菌酶；4，T7 引发酶；5，DNA 聚合酶亚基；5.7，负责在λ感染的细胞中生长；6，外切核酸酶。第 III 组基因（主要与噬菌体粒子的结构和装配有关）：7，7.3，寄主范围的选择；8，头-尾蛋白；9，头部装配；10，壳体蛋白；11、12，尾部蛋白；13、14、15、16，核心蛋白；17，尾丝蛋白；18，19，DNA 成熟有关的蛋白。

## 四、T7 噬菌体生长过程中用噬菌体 RNA 聚合酶代替寄主的 RNA 聚合酶

T7 噬菌体是 *E.coli* 的烈性噬菌体，其 DNA 含有39936bp(不到 T4 的四分之一)，共编码约50 个基因。图 8-19 为 T7 噬菌体的遗传图。

T7 噬菌体在 30℃ 时生长周期大约为 25 分钟(见表 8-5)。从 T7 的生长周期可以看出，其突出的特征是其 DNA 注射非常缓慢，大约经过 10 分钟才完成。而其他噬菌体都是在一分钟以内完成这一过程。这是 T7 噬菌体基因调控的手段之一，因为其基因进入寄主细胞之前是不可能被转录的。

按 T7 的生长周期，可以分为早期转录阶段和晚期转录阶段。早期转录使用 *E.coli* 的 RNA 聚合酶。晚期使用 T7 基因 1 所编码的 RNA 聚合酶。这是 T7 噬菌体实现早晚期基因表达控制的关键。

早期 mRNA 为转录物 I，晚期 mRNA 为转录物 II 和 III。转录物 I，II，III 中，每一组内的各 mRNA 从几个不同的启动子开始转录，而终止点通常是相同的 ($T_E$，$T_\phi$ 和 T7 末端)。

表 8-5　T7 噬菌体的生长周期(30°C)

| 时　间 (分) | 主　要　事　件 |
|---|---|
| 0 | 噬菌体吸附于寄主表面 |
| 0～1 | 噬菌体 DNA 开始向寄主细胞内注射 |
| 2 | 噬菌体 mRNA 合成开始 |
| 4 | 开始关闭寄主基因的转录 |
| 8～9 | 开始合成噬菌体 DNA |
| 8～10 | 结构蛋白质开始合成；噬菌体 DNA 注射完毕 |
| 15 | 第一个子代噬菌体出现 |
| 25 | 寄主细胞溶解，子代噬菌体释放 |

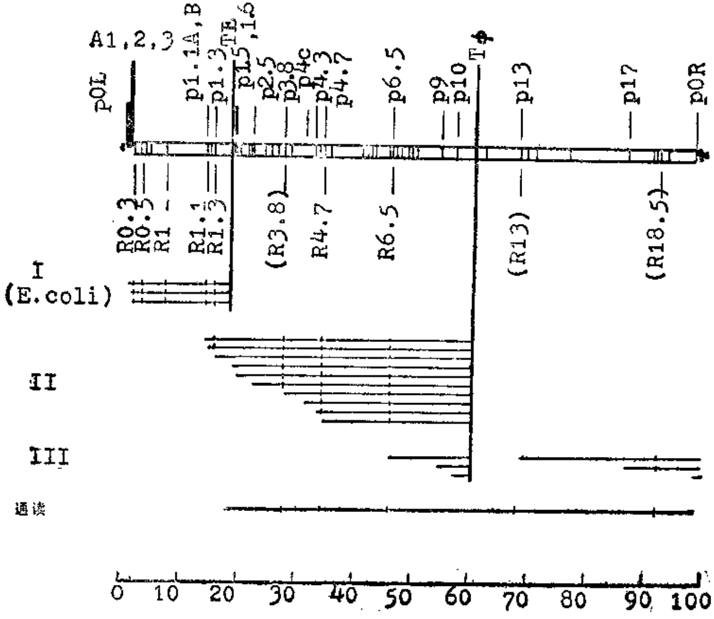



图 8-20　T7 mRNA 的合成和转录后处理。T7 基因用小方块表示，基因上方所示为启动子位置，而下面表示 RNAase Ⅲ 的切点(其中 R3.8 和 R13 并不在每个 mRNA 上都能产生切点)。三组转录物分别用三组水平线表示。下面为越过 Tφ 的通读 mRNA，实际也为一组产物。

图 8-20 表示出这三组转录物的起点和终点。转录物 I 包含 10 个基因，其中最为重要的是基因 0.3，0.7，和 1。基因 0.3 的产物能抑制寄主的限制酶。基因 0.7 的产物是一种蛋白质激酶，能将 *E.coli* 的 RNA 聚合酶磷酸化而使之失活。如果基因 0.7 突变，则噬菌体爆炸规模则显著减小，因为寄主可以继续合成自己的 RNA 而浪费了资源。基因 1 的产物就是 T7RNA 聚合酶，是一条单肽链，分子量 98KDa。它与 *E.coli* RNA 聚合酶比较起来简单得多，因为它所识别的启动子是一个高度保守的 23bp 的序列(－17～＋6)。第 Ⅲ 组五个启动子和 φOR 为

较强的启动子，而第Ⅱ组 10 个启动子和 φOL 为相对弱的启动子，因为其中有 2～7 个碱基不同于一致序列(表 8-6)。

**表 8-6 T7 RNA 聚合酶的启动子**

| 启动子 | RNA 起点 核苷酸 | RNA 起点 T7单位 | 核苷酸序列 |
|---|---|---|---|
| | | | 保守序列 |
| | | | −10 1 |
| | | | TAATACGACTCACTATAGGGAGA |
| | | | 复制起动子 |
| | | | −20 −10 1 |
| φOL | 405 | 1.01 | TTGTCTTTAT TAATACAACTCACTATAAGGAGA GA |
| | | | 第Ⅱ组启动子 |
| | | | −20 −10 1 |
| φ1.1A | 5,848 | 14.64 | AACGCCAAAT CAATACGACTCACTATAGAGGGA CA |
| φ1.1B | 5,923 | 14.83 | TTCTTCCGGT TAATACGACTCACTATAGGAGAA CC |
| φ1.3 | 6,409 | 16.05 | GGACTGGAAG TAATACGACTCAGTATAGGGACA AT |
| φ1.5 | 7,778 | 19.48 | AGTTAACTGG TAATACGACTCACTAAAGGAGGT AC |
| φ1.6 | 7,895 | 19.77 | TGGTCACGCT TAATACGACTCACTAAAGGAGAC AC |
| φ2.5 | 9,107 | 22.80 | AGCACCGAAG TAATACGACTCACTATTAGGGAA GA |
| φ3.8 | 11,180 | 27.99 | CGTGGATAAT TAATTGAACTCACTAAAGGGAGA CC |
| φ4C | 12,671 | 31.73 | CCGACTGAGA CAATCCGACTCACTAAAGAGAGA GA |
| φ4.3 | 13,341 | 33.41 | AGTCCCATTC TAATACGACTCACTAAAGGAGAC AC |
| φ4.7 | 13,915 | 34.84 | TTCATGAATA CTATTCGACTCACTATAGGAGAT AT |
| | | | 第Ⅲ组启动子 |
| | | | −20 −10 1 |
| φ6.5 | 18,544 | 46.43 | GTCCCTAAAT TAATACGACTCACTATAGGGAGA TA |
| φ9 | 21,864 | 54.75 | GCCGCGAATT TAATACGACTCACTATAGGGAGA CC |
| φ10 | 22,903 | 57.35 | ACTTCGAAAT TAATACGACTCACTATAGGGAGA CC |
| φ13 | 27,273 | 68.29 | GGCTCGAAAT TAATACGACTCACTATAGGGAGA AC |
| φ17 | 34,565 | 86.55 | GCGTAGGAAA TAATACGACTCACTATAGGGAGA GG |
| | | | 复制启动子 |
| | | | −20 −10 1 |
| φOR | 39,228 | 98.23 | CACGATAAAT TAATACGACTCACTATAGGGAGA GG |

转录物Ⅱ和转录物Ⅲ有共同的终止位点(Tφ)。这一终止子的终止效率为 90%，还有 10% 可延长至 T7 末端。Tφ 位于基因 10 和基因 11 之间。基因 10 产物为主要的头部蛋白质，需要量大；而基因 10 后面的基因编码少数头部蛋白质和尾部蛋白质，需要量小，因此 Tφ 具有调节作用。虽然Ⅲ类启动子强度大于Ⅱ类启动子，但Ⅱ类启动子先进入细胞，因此先转录。在Ⅱ类转录物中，对于调控最重要的是基因 2 产物，它是 *E.coli* RNA 聚合酶抑制剂。经过基因 0.7 产物将寄主 RNA 的聚合酶磷酸化，再由基因 2 产物的抑制作用，因此在感染后 8 分钟，*E.coli* 的 RNA 合成体系全部被接管，因为噬菌体已不再需要转录物Ⅰ了。转录物Ⅱ的蛋白质大多为 DNA 复制所必需的酶(见第三章)，也有几种结构蛋白质。由于负责 DNA 装配的蛋白质(基因 18，19)翻译很迟，故虽有结构蛋白质，但不会被过早地装配成噬菌体颗粒。

在感染后15分钟停止合成转录物Ⅱ。这时出现整个转录水平的下降。这可能与基因3.5产物N-乙酰胞壁酰基-L-丙氨酰氨酶有关，这个酶同时也负责细胞的裂解。这个阶段，只有较强的第Ⅲ类启动子才能起始转录。

由以上可以看出，T7首先利用寄主的RNA聚合酶来转录产生自已的RNA聚合酶以及蛋白质激酶，自己的RNA聚合酶又用来转录产生寄主RNA聚合酶的抑制剂，从而完全关闭了寄主的转录体系，建立了自己的转录体系。为了实现这一战略目标，SPO1，T4，T7采用了不同的战术。这些事例都说明转录水平的调控是最主要的和最重要的调控机制。

### 五、λ噬菌体早期、晚期基因转录以及裂解生长和溶原生长的调控

λ噬菌体与前面所讲的SPO1，T4，T7三种噬菌体不同，它是*E.coli*的温和噬菌体，在其生活史中有溶原周期和裂解周期。既然有溶原周期，λ噬菌体就不能采取迅速接管寄主的转录机构并代之以自已的RNA聚合酶的办法，否则就切断了自己的后路(溶原周期)。在λ的整个生命过程中都离不开寄主的RNA聚合酶，λ噬菌体的抗终止因子N和Q以及调节蛋白质Cro，CI，CⅡ等均不永久性地改变RNA聚合酶的构象及其对启动子的识别能力。λ噬菌体通过极其微妙的机制选择裂解周期或溶原周期，也就是说要经过"深思熟虑"，再加上其他许多原因，使得λ的生活周期比前述三种烈性噬菌体的生活周期长得多，大约为45分钟。

λ噬菌体基因转录过程中的终止及抗终止机制已在第六章中讲过；与裂解生长有关的DNA复制已在第三章中讲过；作为溶原化过程不可缺少的部分即λ噬菌体与寄主染色体之间的位点特异性重组(整合)将在第十一章讲解。这一小节我们将以λ噬菌体裂解生长和溶原生长的选择为中心讲解λ噬菌体基因表达的调控。其中两种调节蛋白质CI和Cro作用的分子机制以及原噬菌体的诱导将在第十章中讲授。

当λ噬菌体侵染并进入*E.coli*细胞后，线形的DNA粘性末端就相互配对并由DNA连接酶连接成封闭环，再由DNA旋转酶使之转变为负超螺旋的共价封闭环。这一过程在侵染后3～6分钟内完成。

当λ噬菌体DNA进入寄主细胞后，DNA上没有任何调节蛋白质，因此寄主的RNA聚合酶便分别结合于$P_L$和$P_R$。由$P_L$转录出mRNA L1并产生抗终止蛋白N，由$P_R$转录出mRNA R1并产生调节蛋白质Cro。由于N蛋白质的抗终止作用，使得RNA聚合酶能越过终止子$t_{L1}$和$t_{R1}$，$t_{R2}$，从而转录出mRNA $L_2$和mRNA $R_2$，$R_3$；然后分别翻译出CⅢ蛋白，CⅡ蛋白以及O蛋白、P蛋白。由CⅡ和CⅢ共同作用，在*cro*基因和CⅡ基因之间的启动子$P_E$(E为确立之意，establishment)确立了CI基因的转录。从$P_E$向左转录的mRNA中除了CI mRNA外，还包括*cro*基因的反义RNA。这段反义RNA是否能抑制*cro*mRNA的翻译现在还不清楚，但看来是有可能的。由于从$P_E$转录的CI mRNA上含有S-D序列，因此具有很高的翻译活性。现在还不知道从$P_R$由左向右转录的RNA聚合酶是否会与从$P_E$由右向左转录的RNA聚合酶相撞。可能由于$P_E$在CⅡ，CⅢ蛋白的帮助下要比$P_R$强得多因此对$P_E$的利用有可能抑制了对$p_R$的利用。

CI和Cro是λ噬菌体的两种调节蛋白质，它们都结合于左右两个操作子上，只是与每个操作子中的不同位点的亲和力次序不同。CI蛋白与操作子的亲和力次序是$O_{L1}>O_{L2}>O_{L3}$，$O_{R1}>O_{R2}>O_{R3}$；而对于Cro蛋白来说，则$O_{L3}>O_{L2}>O_{L1}$，$O_{R3}>O_{R2}>O_{R1}$(图8-21)。因此，λ噬菌体在侵染寄主后，到底是进行裂解生长，还是进入溶原状态，是要经过一番"斟酌"的，这

个过程就是 CI 蛋白与 Cro 蛋白互相竞争操作子的过程。在这个竞争过程中，CI 蛋白和 Cro 蛋白的数量是决定二者胜负的主要因素，也就是说，看这两个基因谁转录得快，翻译得多。

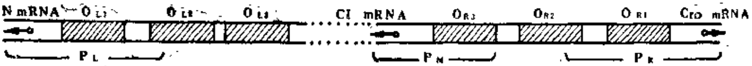



图 8-21 λ 噬菌体操作子 $O_L$ 和 $O_R$ 与启动子 $P_L$, $P_R$ 和 $P_M$。箭头分别表示 N mRNA, Cro mRNA 和 CI mRNA 及其转录起始点。

从时间顺序上，*cro* 基因表达在先，并远早于 CI 基因的表达。而且，如果没有 CⅡ 蛋白和 CⅢ 蛋白的外来帮助，*cI* 基因根本就不能得到表达。由此看来，CⅡ，CⅢ 蛋白又成为 CI 蛋白和 Cro 蛋白谁先占上风的主要因素。在正常情况下，寄主的 *hfl* 基因编码一种蛋白酶，能够水解 CⅡ 蛋白(对于这种机制的拮抗机制留待后面再叙述)。因而，当强大的 *hfl* 基因产物存在时，*cI* 基因是不可能表达的，只有 Cro 遥遥领先，占据了左右两个早期操纵元的操作子。起初允许这两个操纵元表达，后来便封闭这两个操纵元。这时，λ 噬菌体已经产生了足够的与 DNA 复制有关的蛋白质(O, P)和抗终止蛋白(Q)。由于抗终止蛋白 Q 的作用，噬菌体的结构蛋白质基因(A……J)和溶菌酶基因(S, R, $R_z$)就得以表达(见图 2-7)。然后，由基因 A 和 Nul 等产物将环形 DNA 在 cos 位点切断而包装到噬菌体颗粒中，最后溶菌而释放。λ 噬菌体的典型爆炸规模为每个细胞 100 个后代噬菌体颗粒。

在什么样的情况下，λ 噬菌体会走上溶原化的道路呢？就目前所知，有两种因素会导致 λ 噬菌体进入溶原状态：

1. *营养耗竭* 当寄主能源濒于枯竭时，寄主细胞内产生了一系列的生理变化，其中重要的要算是cAMP的产生。目前人们并未发现cAMP-CAP对λ噬菌体的基因表达有调控作用。人们认为，寄主拮抗溶原化的功能即*hfl*基因的表达受到cAMP(可能仍然通过 cAMP-CAP的形式)的负调控。这样，分解 CⅡ 蛋白的蛋白水解酶大大减少，而 CⅢ 蛋白又能够抑制这种蛋白水解酶。寄主的另一个基因 *him*A 也受到细胞营养状态的调控，可能是受 cAMP-CAP 的正调控。*himA* 基因产物是 λ 噬菌体与寄主染色体整合所不可缺少的（作为整合酶寄主因子的一个亚基)。由此看来，λ 噬菌体的整合可能是斗争双方妥协的结果。

2. MOI 值过高（≥10） MOI 值是指感染时 λ 噬菌体与细菌的比例，又称为感染复数(multiplicity of infection)。MOI 值过高这一因素是通过影响 CⅡ 蛋白的活性起作用的。CⅡ 蛋白单体是没有活性的，而且单体很不稳定；只有 CⅡ 的寡聚体才有活性。因此，当一两个 λ 噬菌体同时感染一个寄主细胞时，其 CⅡ 产物不足以形成有活性的寡聚体。这时，λ噬菌体就进入裂解生长。当 MOI 值过高时，就有足够的 CⅡ 寡聚体产生。CⅡ 寡聚体的增加，*hfl* 基因产物的减少，再加上 CⅢ 蛋白协助(抑制 *hfl* 基因产物的蛋白水解酶活性)，CⅡ 蛋白就将从 $P_E$ 转录 *cI* 基因的体制确立起来。由于从 $P_E$ 转录的 CI mRNA 具有很高的翻译活性(比从 $P_M$ 转录的 CI mRNA 翻译活性高 7～8 倍)，因此，CI 蛋白的数量很快地占了上风，而且 $P_E$ 的活跃转录本身就抑制了 *cro* 基因的表达。因而 CI 蛋白就迅速地占领两个早期操纵元的操作子。这样，两个早期操纵元就被封闭，CⅡ，CⅢ 基因也就不再表达。已有的 CⅡ 蛋白和 CⅢ蛋白也很快分解。$P_E$ 也就不能再启动 *cI* 基因的转录。然而，*cI* 基因还有另一个启动子，即 $P_M$，M为维持的意思(maintenance)。由于CI 蛋白的正调控作用，使RNA聚合酶能够从$P_M$转录

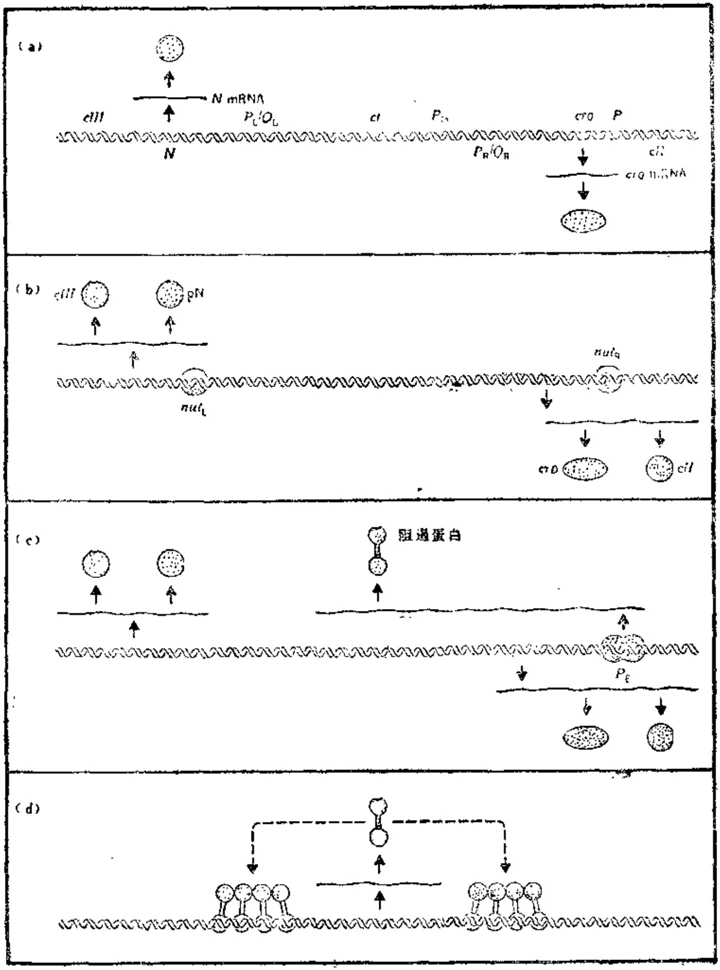



图 3-22　溶原状态的建立需要连瀑式的调控体系，但这种体系最后被阻遏蛋白关闭，代之以自动调节的阻遏蛋白维持系统。

(a) N 和 cro 基因转录。(b) pN 抗终止，因此 CⅡ，CⅢ 基因得以转录。(c) CⅡ 作用于 $P_E$，起始阻遏蛋白基因的转录。(d) 阻遏蛋白结合于 $O_L$ 和 $O_R$，阻断从 $P_L$ 和 $P_R$ 的转录；在 $P_M$ 建立阻遏蛋白基因的自控表达体系；CⅡ/CⅢ 蛋白分解，$P_E$ 不再使用。

*cI* 基因。从 $P_M$ 起始转录的 CI mRNA 没有任何前导序列，缺少核糖体的结合位点(S-D 序列)，因而翻译效率很低。然而这样的数量也就足以维持溶原状态(图 8-22)。在一般情况下，$O_{R3}$ 是空着的。如果万一 CI 蛋白积累，CI 蛋白就会结合到亲和力较小的 $O_{R3}$ 上，从而暂时阻遏了 CI 基因的表达；否则，将会影响 λ 噬菌体的灵活应变能力。

CⅡ 蛋白除了促使从 $P_E$ 转录 *cI* 基因外，同时还能够促使从 $P_I$ 转录整合酶 (integrase) 的基因。而在缺少 CⅡ 蛋白时，这两个启动子都不能起始转录。这两个启动子都没有与一致序列相近的 −10序列和 −35序列，因此单靠 RNA 聚合酶是不能转录的。离体转录试验表明，如果体系中只含有 $P_E$($P_I$) 和 RNA 聚合酶，则转录不能进行，而加入 CⅡ 蛋白，则能正常转录。CⅡ 结合于 −45～−25，加入 RNA 聚合酶后，−12～+11 也被保护。通过对启动子 $P_E$ 的 $C_y$ 突变体(具有与 CⅡ，CⅢ 相似的效应即不能确立溶原状态)的分析，发现突变体可分为两组：一组为 $C_yL$，位于 −10 序列，这类突变体不能结合 RNA 聚合酶；另一组为 $C_yR$，位于 −35 序列。$C_yR$ 突变体又有两类：一类位于经典的 −35 序列位置，其突变体不影响 CⅡ 的结合，而影响 RNA 聚合酶的结合；一类突变位于上述位置的两翼，构成各有四个碱基对的正向重复序列：$\begin{matrix}\text{TTGCGTTTGTTTGC}\\\text{AACGCAAACAAACG}\end{matrix}$。其中相应的重复部分相距 10bp，因此 CⅡ 蛋白（可能为二聚体）结合于 DNA 双螺旋的同一侧。这样的正向重复序列发生突变，$P_E$ 就失去了结合 CⅡ 蛋白的能力。

尽管左向早期操纵元的转录产物 $L_2$ mRNA中包含了整合酶的基因 *int*，但这个基因的信息不能表达成蛋白质(其机制将在本章第七节中讲解)，因为进行裂解生长，是用不着整合酶的。然而，整合过程又必须有整合酶的存在。因此从 $P_I$ 转录基因 *int*，是建立溶原状态必不可少的一环。由于 $P_I$ 位于 *xis* 基因内，因此从 $P_I$ 的转录活动则排除了 *xis* 基因表达的可能性，这就保证了溶原状态的稳固建立。溶原化过程是λ噬菌体的 *att* 位点与寄主的 *att* 位点进行位点特异性重组的过程，其详细机制留待第十一章中叙述。

CⅡ 蛋白除了作为正控制因子促进 $P_E$ 和 $P_I$ 的转录活性外，还能阻碍 Q 基因的表达，并由此阻碍晚期基因的转录，其详细机制将在第七节中叙述。至于 CⅡ 在裂解生长中的作用机制，仍毫无所知。

λ 噬菌体的裂解生长和溶原生长的调控是非常细腻的，以致于将野生型λ噬菌体接种于同一块培养基的细菌中，有的出现裂解生长，形成了噬菌斑，而又有些噬菌体与寄主整合后生成的溶原菌再次生长出来，出现在原来的噬菌斑中，使噬菌斑变得模糊不清。这种情况下出现两种不同的生长可能是由于 MOI 值的不同造成的。如果 CI 基因发生突变，则 λ 噬菌体不能进入溶原周期，只得到清亮的(clear)噬菌斑，*cI* 基因便由此得名。如果操作子发生突变，便不能结合 CI 蛋白，λ 噬菌体同样不能进行溶原生长，而只能进行裂解生长，因此这种突变又称为烈性突变 (virulent mutation)，实质上相当于细菌的 $O^C$ 突变。后来发现，$c$Ⅱ$^-$，$c$Ⅲ$^-$ $cy^-$ 等突变体也都能使 λ 噬菌体产生清亮的噬菌斑。但不同的是，突变体 $cI^-$ 既不能建立溶原状态又不能维持溶原状态，而突变体 $c$Ⅱ$^-$，$c$Ⅲ$^-$，$cy^-$ 等虽然难以建立溶原状态，而一旦建立后便能像野生型 λ 噬菌体那样维持溶原状态。因此，$c$Ⅱ，$c$Ⅲ 基因以及 *cy* 位点 ($P_E$) 与建立溶原状态时 λ 阻遏蛋白的合成有关，而溶原状态建立之后就不再需要它们了。

# 第六节 基因转录的翻译调控——衰减子系统[39~43]

## 一、概 述

在色氨酸操纵元中，阻遏状态和除阻遏状态使从启动子开始的转录起始频率相差大约70倍，也就是说，本底合成是除阻遏合成的1/70。这与*lac*操纵元中的情况相比，其阻遏效率要低得多。在*lac*操纵元中，阻遏状态下的合成仅为诱导状态下合成的1/1000。然而在阻遏状态下，trp酶系统的活性(酶的浓度)却只有除阻遏状态下的1/700。这就是说，在色氨酸操纵元内必然还存在着另一个控制基因表达的系统。

前面我们所讲过的阻遏-除阻遏的调控系统或多或少都带有全或无的性质。而对于许多氨基酸合成的操纵元，需要对体内的氨基酸水平作较为细致的调节(其生物学意义还不完全清楚)，因此便演化出更为精细的调节系统。色氨酸操纵元的衰减子系统就是一个典型的例子。

衰减子系统的控制和阻遏蛋白的控制在方向上是相同的。在色氨酸存在时，两种系统都具有使转录水平降低的功能。衰减作用的幅度大约为10倍。在无色氨酸存在时，所有转录trp操纵元的RNA聚合酶都可以通过衰减子。当有色氨酸存在时，一部分RNA聚合酶逃过阻遏蛋白的监督而起始转录；而这些RNA聚合酶在衰减子处大部分被扣留，而只有大约10%的RNA聚合酶能侥幸通过。这样，使*trp*操纵元在缺乏色氨酸以及含有不同浓度的色氨酸时能在大约700倍的范围内进行调节($70\times10=700$)。

## 二、衰减子的作用机制

衰减子是如何控制trp操纵元的转录的呢？通过对*trp* mRNA上前导序列的分析，现在已基本上搞清其作用机制。在*trp*E上游的mRNA前导序列中有一个编码14个氨基酸的开放阅读框架，其前面并有一个核糖体结合位点(S-D序列)。在细胞内虽然未分离到这样的14肽，但将其S-D序列与结构基因相融合，能进行有效的翻译。开放阅读框架下游相隔42个碱基，就是一个典型的不依赖$\rho$因子的终止子(28bp)。这个终止子就是衰减子的核心部分，有时为了简便起见，就把这个部分称为衰减子序列(图8-23)。

衰减子序列本身并不能实现衰减作用，而必须通过对前导序列上14个氨基酸的前导肽(leader peptide)的翻译才能实现。因此衰减作用的实质是以翻译手段控制基因的转录。前导肽mRNA上有两个连续的色氨酸密码子。在大肠杆菌蛋白质中，色氨酸是很少出现的氨基酸。因此前导序列中两个连续的色氨酸密码子的出现并非巧合，而是有其特殊意义的。人们早就发现，色氨酰-tRNA合成酶的突变使得*trp*操纵元不能达到最大程度的阻遏。这类突变影响到细胞内色氨酰-tRNA的合成，当核糖体进展到两个连续的色氨酸密码子时由于缺少色氨酰-tRNA而陷于停顿。色氨酸的缺乏也造成同样的后果。就是这种核糖体的停顿，才使聚合酶RNA越过衰减子的障碍，继续完成*trp* mRNA的转录。

RNA聚合酶是在衰减子序列上终止转录还是继续转录下去，是通过*trp* mRNA的前导序列上能形成两种不同的二级结构所决定的。在图8-24中，中图为*trp* mRNA中能形成碱基配对四段序列的模式图，其中序列1能与序列2配对，序列2又能与序列3配对，序列3又能与序列4配对。左图表示序列1与序列2配对($\Delta G=-11.2$Kcal/mole)；序列3与序列4配对($\Delta G=-20$ Kcal/mole)。右图表示序列2与序列3配对($\Delta G=-11.7$ Kcal/mole)。

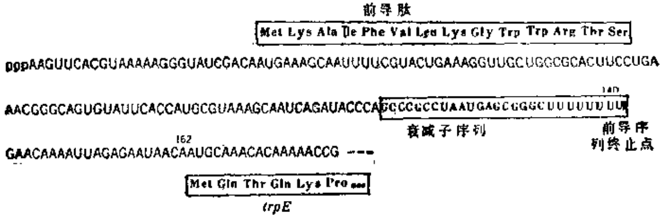



图 8-23 *trp* 操纵元 mRNA 的前导序列

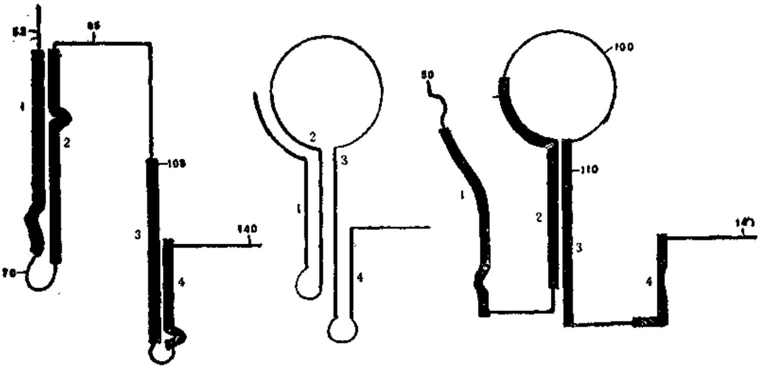



图 8-24 trp mRNA 的前导序列可以形成两种可能的二级结构

体外实验中配对形式如左图所示，而并未发现右图形式。然而在细胞内，当序列 2 和序列 3 形成双链结构时，序列 4 尚未转录完毕。因此只要序列 1 被核糖体占据，序列 2 和序列 3 就可能形成双链结构。

在细菌中，由于没有核膜的阻隔，因而转录和翻译是紧密耦联的。可能 RNA 聚合酶刚转录出前导肽的部分密码子，核糖体就开始翻译。当细胞中有色氨酸存在时，核糖体能够顺利地翻译出整个前导肽而在终止码密子 UGA(+70) 处停下来。这时核糖体占据了序列 1 和部分序列 2，使序列 2 和序列 3 不能产生有效的配对，因而序列 3 和序列 4 配对产生终止子的发夹结构，于是实现转录的终止(图 8-25 a)。当出现 Trp (或 Arg)饥饿时，核糖体停顿在两个 Trp 密码子上。这时，核糖体占据了序列 1，而留下完整的序列 2 以便与转录出的或即将转录出来的序列 3 形成二级结构。这样当序列 4 转录出来后仍然是单链状态，即终止子不能形成，于是转录继续进行下去 (图 8-25 b)。其他的氨基酸饥饿，都不能产生图 8-25 b 的情况。例如当 Gly 饥饿时，核糖体停顿在 +52 的 Gly 密码子上。这时，还保留了序列 1 与序列 2 的部分二级结构，序列 2 和序列 3 之间就不能形成二级结构，而即将转录出来的序列 4 就能与序列 3 形成二级结构，使转录终止(图 8-25 c)。如果没有翻译作用存在，则会产生图 8-24 左图所示的二级结构，则转录必定在衰减子处终止。

上述模型得到了大量实验证据的支持。影响序列 3 和 4 二级结构的稳定性的突变和改变

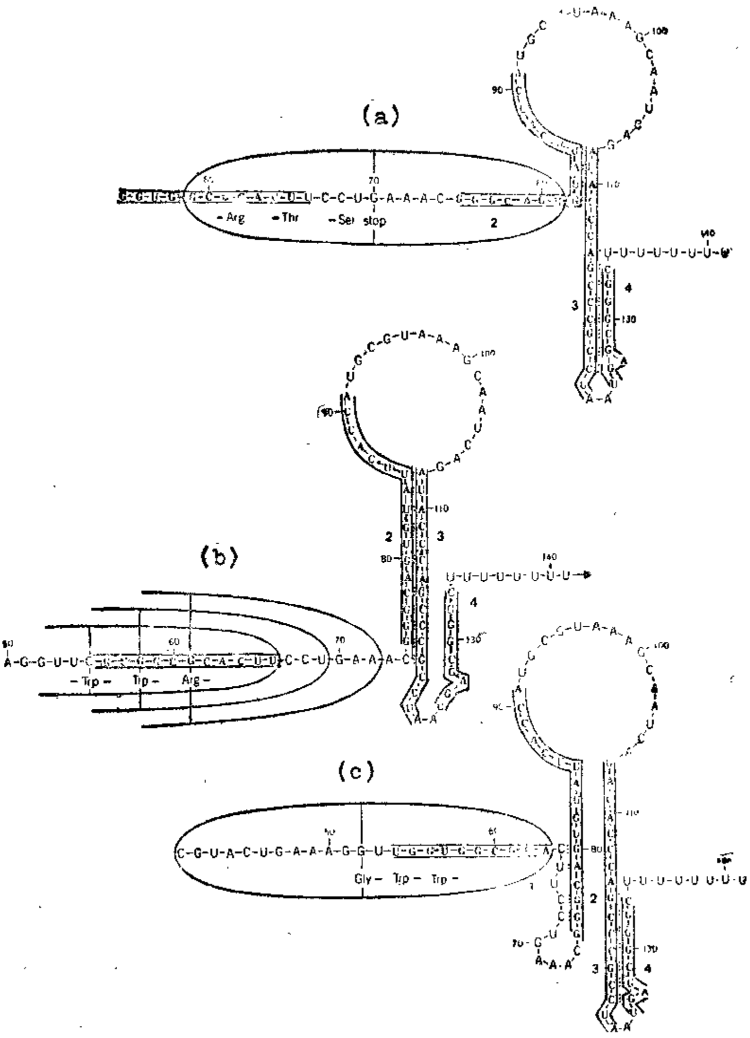



图8-25　核糖体的位置控制着 *trp* 前导序列 mRNA 的构象，几个碱基的偏差就能决定序列 3 和 4 是否形成终止子发夹结构

(a) 氨基酸供应充足时，转录终止；　(b) 色氨酸或精氨酸饥饿时，转录继续进行；　(c) 甘氨酸饥饿时，转录亦终止

序列 4 后面一连串 U 的突变均降低了衰减子的转录终止作用而增加了 Trp 酶系统的表达，这与通常的终止子突变效应是一样的。相反，影响到序列 2 和 3 配对的碱基突变，则增加衰减子的衰减作用。如果前导肽的起始密码子突变成为别的碱基序列，则衰减子只能起终止子的作用，从而极大地降低了 Trp 酶系统的表达。

由上述可见，实现衰减子对转录的调控的关键在于空间上和时间上的巧妙安排。在空间上，两个色氨酸密码子的位置至关重要，不可随意变更，否则就不能实现衰减。在时间上，核糖体停顿于两个色氨酸密码子上时，序列 4 应当还未转录出来，否则就不会出序列 2 和 3 的配对，而只能是序列 3 和 4 的配对，也就无法使转录越过衰减子而继续下去。这种时间上的巧妙安排的一个原因就是 RNA 聚合酶在转录完 +90 时产生一次延宕。这就给予核糖体以追赶的机会。

### 三、衰减子的普遍性及其生物学意义

除了 *trp* 操纵元外，其他许多负责氨基酸合成的操纵元都受到衰减子的调控，特别是在 *his* 操纵元中，衰减子是唯一的控制机构。在这些操纵元中，也都存在着前导肽的开放阅读框架，也具有不依赖 $\rho$ 因子的终止子序列，而且在前导肽的 mRNA 中，具有比 *trp* 操纵元前导肽多得多的连续的相应氨基酸的密码子。例如，在 *his* 操纵元的 16 个氨基酸组成的前导肽中，有连续 7 个组氨酸；在 *phe* 操纵元的 15 个氨基酸组成的前导肽中，也有 7 个苯丙氨酸。在某些操纵元中，衰减子可以对不止一种氨基酸饥饿作出反应。例如在 *thr* 操纵元中的苏氨酸和异亮氨酸；*ilv* 操纵元中的异亮氨酸、缬氨酸和亮氨酸(表 8-7)。在这些操纵元中，由于能够引起核糖体停顿的密码子较多，因此可能需要一个以上的核糖体停顿在这一连串的密码子处才能实现有效的除阻遏。这些操纵元的前导序列中，也都可能存在着两种可交替的二级结构，用以调节终止子发夹结构是否形成。同样，当前导肽翻译作用不存在时，则 mRNA 必定在衰减子处终止。

**表 8-7 已知的氨基酸合成操纵元中前导肽的氨基酸序列及其包含的调控氨基酸**

| 操纵元 | 前导肽序列 | 调控氨基酸 |
|---|---|---|
| *trp* | Met Lys Ala Ile Phe Val Leu Lys Gly Trp Trp Arg Thr Ser | Trp |
| *his* | Met Thr Arg Val Gln Phe Lys His His His His His His His Pro Asp | His |
| *pheA* | Met Lys His Ile Pro Phe Phe Phe Ala Phe Phe Phe Thr Phe Pro | Phe |
| *leu* | Met Ser His Ile Val Arg Phe Thr Gly Leu Leu Leu Leu Asn Ala Phe Ile val Arg Gly Arg Pro Val Gly Gly Ile Gln His | Leu |
| *thr* | Met Lys Arg Ile Ser Thr Thr Ile Thr Thr Thr Ile Thr Ile Thr Thr Gly Asn Gly Ala Gly | Thr, Ile |
| *ilv* | Met Thr Ala Leu Leu Arg Val Ile Ser Leu Val Val Ile Ser Val Val Val Ile Ile Ile Pro Pro Cys Gly Ala Ala Leu Gly Arg Gly Lys Ala | Ile, Val, Leu |

一般说来，衰减子的作用机制可以看成是这样一种调控机制：其中核糖体的功能相当于一个正调控蛋白，它在停顿位点的结合是 RNA 继续合成（防止终止作用）所必须的；而相应的氨酰基-tRNA 与之结合则使核糖体不能与停顿位点结合。这个氨酰基-tRNA 的功能则相当于辅阻遏物。这样，通过前导肽的翻译来控制 mRNA 合成的机制就是一种可阻遏的正控制（图 8-1）。

细菌演化出衰减子调控系统，其生物学意义可能是：(1)活性阻遏物和非活性阻遏物的转变可能较慢，而 tRNA 荷载与否可能更为灵敏；(2)氨基酸的主要用途是合成蛋白质，因而以 tRNA 荷载情况为标准来进行控制可能更为恰当；(3) 为什么大多数这样的操纵元又同时需要阻遏蛋白呢？因为衰减子系统需要先转录出前导肽mRNA，然后再根据前导肽的翻译情况来决定 mRNA 是否继续转录。然而当氨基酸供应充足时，就没有必要通过这样的步骤，而直接关闭 mRNA 的转录活性。也就是说，需要一个决定基础水平的控制系统：当细胞内氨基酸高于某一水平时，可以实现完全的阻遏；而只有低于这一水平时，才需要用衰减子这个细调节旋钮来进行调节。两种调节机制都是为了避免浪费，提高效率。

### 四、负责核苷酸合成的基因的衰减子

新近人们发现，负责合成嘧啶核苷酸的基因也具有衰减子结构。在大肠杆菌中，UTP 的前体 UMP 的合成所需的酶由五个操纵元所编码。这五个操纵元都有自己的衰减子，其中以 UMP 合成的第一步反应所需的天门冬氨酸转氨甲酰酶的操纵元的衰减子作用机制较为清楚。这个操纵元叫做 *pyrBI*；其中 *pyrB* 编码催化亚基，*pyrI* 编码调节亚基，二者构成一种具有别构效应的天门冬氨酸转氨甲酰酶；该酶的底物为天门冬氨酸和氨甲酰磷酸。

在 *pyrBI* 操纵元中，*pyrB* 基因的翻译起始密码子之前有 160 个核苷酸的前导序列。其中，从第 21 核苷酸开始，有一个 44 氨基酸的前导肽编码序列。在第 120 核苷酸附近有一个典型的不依赖于 RhO 因子的终止子结构。这一终止子结构与色氨酸操纵元的衰减子中的终止子结构一样，如果没有前导肽的翻译作用的存在，它只不过是一个永久性的终止子。然而，上述两个衰减子中先导肽的翻译对于转录调控的机制却是根本不同的。在色氨酸操纵元的衰减子结构中，先导肽的终止密码子位于终止子上游方向约 45 核苷酸处；而 *pyrBI* 操纵元中，先导肽的终止密码子却位于终止子下游方向约 25 个核苷酸处。色氨酸操纵元中，是由于缺少色氨酸而使核糖体在连续的两个色氨酸密码子处停顿而影响终止子结构的形成，因而其调控位点是连续的两个色氨酸密码子；而在 *pyrBI* 操纵元中，是由于缺少 UTP 使得 RNA 聚合酶在暂停位点 (pause site) 发生停顿，以致使从事翻译的核糖体能够追赶上 RNA 聚合酶。这种起调控作用的暂停位点是由位于第 60 核苷酸附近的发夹结构和第 85 核苷酸附近的一串 U (7 个 U 被 1 个 C 所隔开) 所组成。这一发夹结构后并没有紧跟一连串的 U，甚至也不富含 A/U，因此不能作为终止子；但这一发夹结构却能使 RNA 聚合酶发生延宕，特别是当细胞内 UTP 含量较低时会在一连串 U 处发生较长时间的停顿，从而给予核糖体以追赶的机会。或许正是由于核糖体与 RNA 聚合酶的某种作用，将嵌合于 RNA 聚合酶内部的发夹结构：出来，从而结束了这种长时间的延宕作用。当这种耦联了核糖体的 RNA 聚合酶向前转录衰减子的终止子序列时，可能由于同样的原因使得终止子发夹结构无法形成。如果设法降低核糖体的翻译速度（如在先导肽的编码序列中换上使用频率低的密码子），则可显著地增加转录的终止。而在细胞内存在着充足的 UTP 时，RNA 聚合酶在转录了暂停位点的发夹序列后虽然出

现短时间的延宕，但是 RNA 聚合酶还是很快地越过这个单一的障碍，继续向下游方向转录过去，只到衰减子的终止子处停止了转录。这时，RNA 聚合酶离开了模板，同时释放出大约 135 核苷酸的前导 RNA 序列。*pyrBI* 操纵元的衰减子结构及其可能的调节机制示于图 8-26。

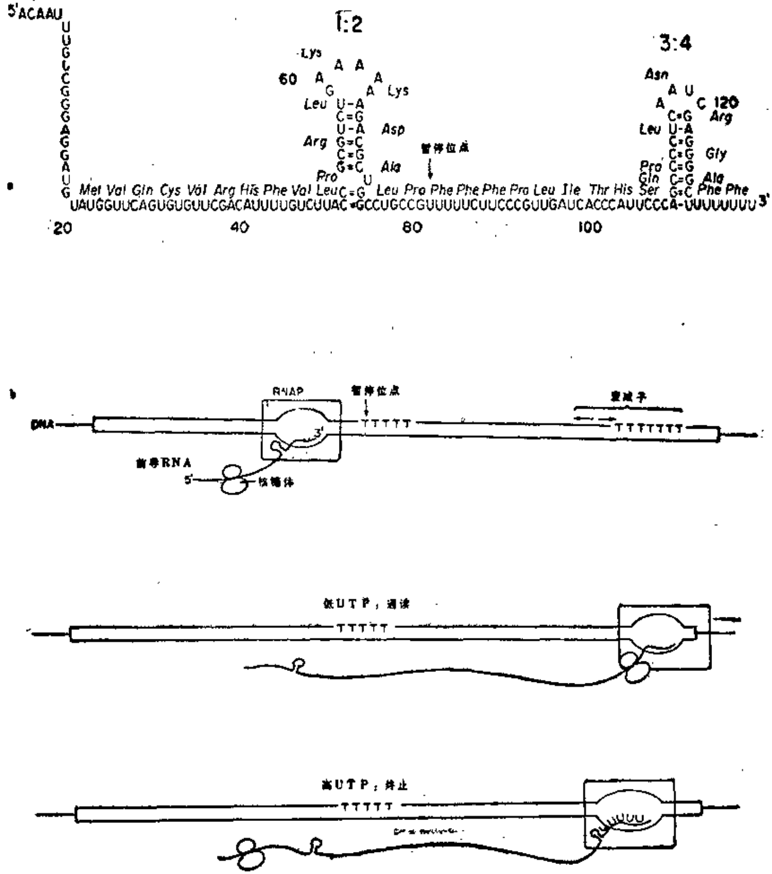


图 8-26 *pyrBI* 操纵元的衰减子结构及其可能的调节机制

(a) 衰减子结构，其中 3:4 为终止子的发夹结构，1:2 为暂停位点的发夹结构；(b) RNA 聚合酶在暂停位点的长时间延宕给予核糖体以追赶的机会；(c) 细胞内 UTP 含量低时，与核糖体耦联的 RNA 聚合酶越过衰减子的终止子序列，产生通读；(d) 细胞内 UTP 含量高时，RNA 聚合酶在衰减子的终止子序列处终止转录，出现衰减。

## 第七节 翻译水平的调控[39,44~67]

### 一、反义 RNA 的调控作用

1983 年，Mizuno，Simon 等人差不多同时发现了反义 RNA 对于基因表达的调控作用，从而揭示了一种新的基因表达调控的机制。在这以前，人们普遍认为基因的调控只有通过蛋白质与核酸的相互作用而实现。而反义 RNA 的调控是核酸与核酸的相互作用。反义 RNA 通过互补的碱基与特定的 mRNA 结合，结合位点通常是 mRNA 上的 S-D 序列，起始密码子 AUG 和部分 N 端的密码子，从而抑制 mRNA 的翻译。人们称这类 RNA 为干扰mRNA 的互补 RNA，简称 micRNA(mRNA-interfering complementary RNA)。

人们最早研究的 micRNA 是调节大肠杆菌外膜蛋白 omFF mRNA 表达的 micF RNA 和转座元 Tn10 中调节转座酶 mRNA(RNA-in)翻译的 RNA-out。关于 RNA-out 将在第十一章讲述。

Mizuno 等人研究了渗透压变化对大肠杆菌外膜蛋白质基因表达的调节，发现了两种外膜蛋白质，一种叫 OmpC，一种叫 OmpF，它们的合成受渗透压的调节。在高渗透压时，OmpC 合成增多，而 OmpF 的合成受到控制，在低渗透压时，OmpF 合成增多，而 OmpC 的合成受到控制。两种蛋白质含量随着渗透压的变化而改变，但两种蛋白质总量保持不变。现在已经知道，当 OmpC 基因转录时，在 OmpC 基因的启动子上游方向有一段 DNA 序列，以相反的方向同时转录产生一个 174 个核苷酸的 RNA，这个 RNA 能够与 OmpF mRNA 的前导序列中的 44 个核苷酸(包括 S-D 序列)以及编码区域(包括起始密码子 AUG)形成杂合双链，从而抑制了 OmpF mRNA 的翻译。OmpC 转录得越多，micF 也就越多，OmpF 蛋白就越少。micF 与 OmpF mRNA 的碱基配对情况示于图 8-27。

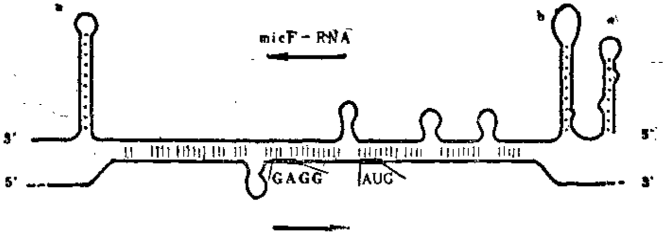



图 8-27 micF 与 OmpF 的 mRNA 所形成杂合双链(除了 S-D 序列和 AUG 之外，其余序列均省略)

人们根据反义 RNA 调控基因表达的原理，使 λ 噬菌体的 CⅡ 蛋白抑制晚期基因转录这个长期令人迷惑不解的问题迎刃而解了。1985 年，Hoopes 等人发现，在 Q 基因的中部有一个依赖于 CⅡ 蛋白的启动子，其序列与 $P_E$ 和 $P_I$ 基本相同。在 −35 序列的两翼，有典型的 CⅡ 蛋白结合位点 TTG NNNNNNNTTGC，其转录的方向与 Q 基因相反。这个启动子就取名为 $P_{aQ}$，即抗 Q 启动子(anti-Q Promter)。其转录产物为一个 ≈200 核苷酸的 RNA，不编码任何蛋白质，其序列与 Q 基因的 5′ 端的一半完全互补，因此有效地抑制了 Q mRNA 的

翻译。同时由于 CⅡ 蛋白的参与，$P_{aQ}$ 与 $P_E$一样，具有很高的转录活性，从而抑制了从 $P_L$ 的转录活性。因此不仅 Q 基因的 mRNA 不能翻译，而且使得 Q 基因的转录活性也降低。如果在 CⅡ 蛋白的结合位点出现一个由 T·A→C·G 的突变，叫做$P_{aq1}$(图 8-28)，则需要增加 4 倍量的 CⅡ 蛋白才能恢复野生型启动子($T_{aQ}$)所具有的转录活性。突变的启动子 $P_{aq1}$ 在低温下(30℃)就能使 λ CI857 突变株产生清亮的噬菌斑而不是浑浊的噬菌斑，因此 $P_{aQ}$ 的转录显然有利于溶原生长。

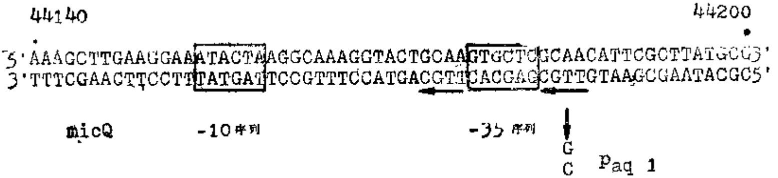



图 8-28 启动子 $P_{aQ}$ 的结构与其突变体 $P_{aq1}$

反义 RNA 的概念可以用于研究基因的功能。传统的方法是使基因的某些碱基发生突变，然后再观察其表现型，但常常有无声突变或渗漏突变产生，影响对突变效应的观察。而反义 RNA 对基因的抑制比较完全，而且是专一性的。只要将一个不含启动子的克隆基因或其片断的 3′ 末端以相反的方向接上一个启动子，然后通过载体引入细胞内，就可以在细胞内产生反义 RNA 以抑制染色体上相应基因的表达。在这方面已有许多成功的实例。反义 RNA 的作用不仅在于抑制核糖体与 mRNA 的结合从而抑制 mRNA 的翻译，在许多情况下还使 mRNA 含量显著减少。因此反义 RNA 的调控机制可能有多种途径，并不仅仅限于翻译水平。

1984 年，Izant 和 Weintraub 提出了用反义 RNA 来治疗病毒所引起的细胞转化。这是非常令人注目的。在治疗病毒疾患时，很难找到专一性的药物，因为病毒在它的生活周期中常使用寄主的生理机能。反义 RNA 的专一性是毫无疑义的，但要有一个合适的抗病毒载体和有效的给药途径，还必须作出巨大的努力。

## 二、mRNA 本身的二级结构影响翻译的进行

mRNA 的二级结构可以在不同的水平上对基因表达实行调控。在转录水平上起作用的如转录的终止子，衰减子；在翻译水平上起作用的如本节所谈到的反义 RNA，其二级结构是生成于两个 RNA 分子之间。这一小节以及下一小节我们要讲的是 mRNA 本身的二级结构对于翻译的影响。

大肠杆菌有几种十分相似的单链 RNA 噬菌体，如 MS-2，R17，f2，Qβ 等。这些噬菌体都是无尾的正 20 面体，其 RNA 为大约 3 600 个核苷酸(前三种)或大约 4 200 个核苷酸 (图 8-29)。每种噬菌体的基因组成都十分类似，各有四个基因，其中三个基因的序列都很类似，从 5′ 端到 3′ 端依次为附着蛋白(A 基因)，衣壳蛋白(CP 基因)，复制酶(*rep* 基因)。在 MS2 家族(MS2，R17，f2)中，A 蛋白有 393 个氨基酸，分子量 44KDa，每个噬菌体粒子中只有一个分子。衣壳蛋白为 129 个氨基酸，分子量 13.7KDa，每个噬菌体粒子中有 $2\times10^6$ 个分子。复制酶蛋白由 544 个氨基酸组成，分子量为 65KDa，它是构成复制酶复合体的亚基之一。而 Qβ 的这三个基因产物的分子量略大一点。第四个基因的位置较为特殊(见图 8-29)，在MS2，

R17, f2 三种噬菌体中，编码由 75 个氨基酸组成的裂解蛋白，基因序列与 *CP* 基因和 *rep* 基因重叠。而 Qβ 的第四个基因为 *CP* 基因的延伸，通过无义抑制 tRNA 在 CP mRNA 上的 UGA 处插入色氨酸，从而得到 38KDa 的次要衣壳蛋白 $A_1$(327 个氨基酸残基)。 $A_1$ 蛋白不具备裂解功能，但它是 Qβ 侵染寄主所不可缺少的蛋白质。$A_2$ 蛋白兼具裂解功能。

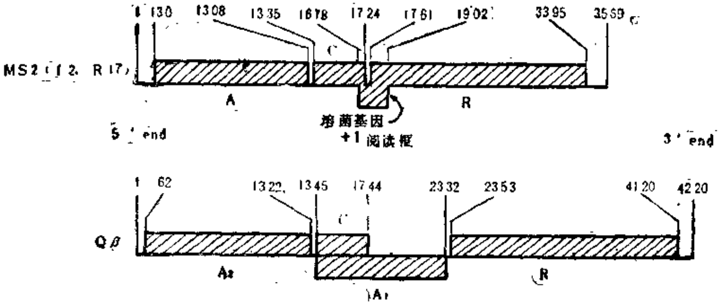



图 8-29 单链 RNA 噬体菌 MS2(f2,R17) 和 Qβ 的遗传图。
斜线部分表示编码序列。数字为 5′ 到 3′ 方向的核苷酸位置。

对于四种基因表达的调控，以 R17 研究得最为清楚。当它的 RNA 进入寄主细胞后，分子内形成许多碱基配对的二级结构，A 基因和 *rep* 基因的核糖体结合位点均处于二级结构中，只有 *CP* 基因的核糖体结合位点游离于二级结构之外，可以为核糖体识别而合成衣壳蛋白。核糖体对 *CP* 基因的翻译冲开了 *rep* 基因核糖体结合位点的二级结构，核糖体才能与之结合翻译出 RNA 复制酶。在 R17 生命过程中，CP 蛋白质的需要量为 $2\times10^6$ 个分子，要比催化 RNA 复制的 RNA 复制酶用量大得多。当 CP 蛋白质达到一定浓度时，它又可以与 *rep* 基因的核糖体结合位点相结合，又封闭了 *rep* 基因。这样就避免了合成过多的 RNA 复制酶而造成的浪费。

RNA 复制酶(分子量 65KDa)合成以后，便与寄主的三个蛋白质分子 Tu(30KDa)，Ts (45KDa)，S1(70KDa) 结合成为 RNA 复制酶复合体。其中噬菌体编码的 RNA 复制酶起催化作用，而 $T_u$ 和 $T_s$ 负责识别正链和负链模板的 3′ 末端从而促成复制的起始。正链和负链的 3′ 末端均具有 tRNA 3′ 末端CCA-OH 的序列，因此 *Tu* 和 *Ts* 对这一序列的识别与它们在蛋白质合成过程的功能相类似。S1 只有在以正链为模板复制负链时是所必须的，这可能是因为正链噬菌体进入细胞后只有这个唯一的模板，S1 能够帮助复制酶复合体识别噬菌体 RNA 上的特定的结合位点。

复制酶复合体组成以后，以原有的噬菌体 RNA(正链)为模板复制出负链，再以负链为模板产生更多的正链。在正链刚开始复制时，与 A 基因核糖体结合位点互补的序列还没有复制出来。这时候，核糖体就乘机结合上去进行翻译，当结合了1～2个核糖体后，互补序列就复制出来了。核糖体再也无法进行翻译活动(图 8-30)。这就是说，每产生一个正链，大约有 1～2个 A 蛋白质合成出来。实际上，每个成熟的噬菌体颗粒只含有一个分子的 A 蛋白，因此这种调控是十分经济而有效的。

mRNA的二级结构不但可以通过影响核糖体的结合而实行调控，而且mRNA的二级结构

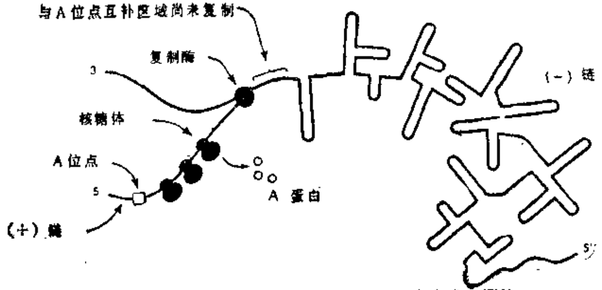



图 8-30 RNA 二级结构对于基因 A 蛋白合成的调控
图中"A 位点"为 A 基因核糖体结合位点的简称。

也是决定 mRNA 寿命的重要因素之一。

### 三、mRNA 的寿命对基因表达的调控

一般说来，原核生物的 mRNA 寿命很短，大多数 mRNA 的半衰期为 2～3 分钟，这样就能够使细菌迅速适应环境的变化。但是细菌中各种 mRNA 在寿命上还是有相当大的差异的。决定 mRNA 寿命的许多因素都影响到基因的表达。

不同的 mRNA 有不同的降解速度。近年来的研究表明，mRNA 的降解作用并不是随机的。如果是由于随机的内切酶作用而降解的话，长的 mRNA 分子应当比短的 mRNA 更不稳定，然而实际上并非如此。Belasco 等人将 *E. coli* 的 *bla* 基因(编码 β-内酰胺酶)与其大片段同框缺失突变体的 mRNA 进行比较，二者的大小相差约 50%，然而它们具有类似的降解速度。

人们很早就提出核糖体对 mRNA 的保护作用，但是近来的研究表明这种作用并不是决定 mRNA 降解的重要因素。Stanssens等人比较了具有不同的核糖体结合能力的各种 *lacZ* 转录物，发现它们具有差不多相同的寿命。

降解 mRNA 的酶主要是 3′-外切核酸酶。而 mRNA 在分子末端的二级结构可能阻止了 3′-外切酶的进攻。人们早就发现不依赖于 ρ 因子的终止子结构使其 mRNA 更为稳定。通过基因融合试验，发现这种终止子序列给融合基因带来更大的稳定性。凡是降低终止子中柄-噜噗结构强度的突变都造成 mRNA 稳定性的降低。由此可见，终止子结构的意义不仅在于转录的终止，而且决定了 mRNA 的寿命。在细胞内，决定 mRNA 稳定性的因素可能更为复杂，因为具有不依赖于 ρ 因子的终止子结构的某些 mRNA 仍然是不稳定的。这就表明，可能存在着具有一定序列专一性的核酸内切酶能够影响 mRNA 的稳定性。最典型的例子就是 λ 噬菌体 *int* 基因的表达。

λ 噬菌体 *int* 基因在左向早期操纵元中。当抗终止蛋白 N 产生之后，左向转录物 $L_2$ 中就包含了 *int* 基因 mRNA。这时，λ 噬菌体尚未决定是否进入溶原周期，因而，λ 噬菌体并不需要整合酶。原来，λ 噬菌体在其生活史的不同时期，能够产生三种不同的但都包含 *int* 基因的 mRNA。由于抗终止蛋白 N 的作用，左向转录物 $L_2$ 就越过了终止子 $t_{L2}$ 直至一个很强的终止子处才终止转录。这样，在已转录的终止子 $t_{L2}$ 序列所形成的柄-噜噗结构上又增添了一个附加的柄-噜噗结构(图 8-31)。而这个附加的柄-噜噗结构恰恰构成 RNAaseⅢ的

识别位点和作用位点（又称为 *sib* 位点），在此处将左向转录物 $L_2$ 切断。这样，整合酶 mRNA 的 3′ 末端就暴露于外切核酸酶的锋芒之下而很快降解。RNAase Ⅲ 基因突变 ($rnc^-$) 或 *sib* 位点的突变 ($sib^-$) 都导致整合酶含量的增加。由于 *sib* 位点位于整合酶基因的下游，而且整合酶 mRNA 的降解是从 3′ 向着 5′ 方向进行，因此 *sib* 位点对于整合酶基因表达的负调控又称为返回调控(retroregulation)。

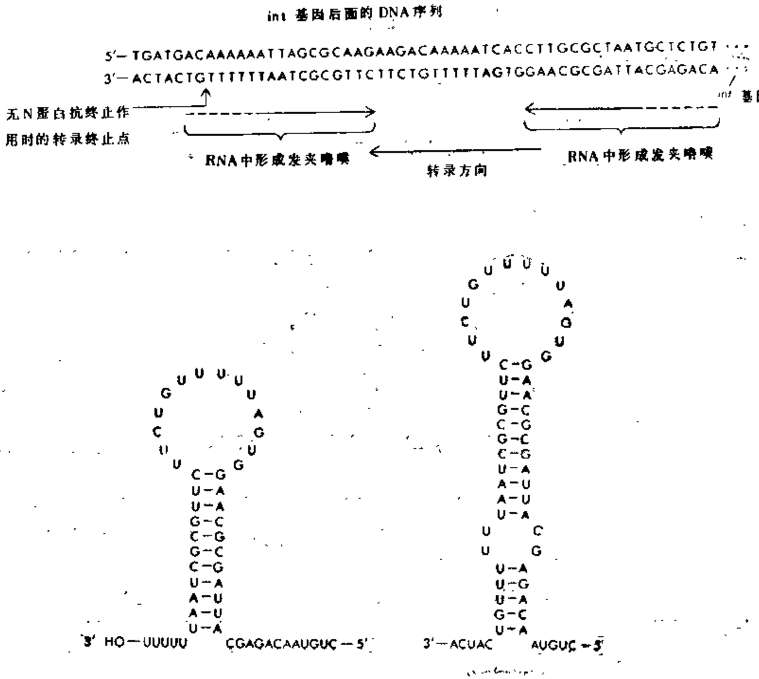



图 8-31 （上）游离的 λ 噬菌体的终止子 $t_{L2}$ 和 Sib 位点的 DNA 序列；
（右下）游离的 λ 噬菌体的左向转录物 $L_2$ 中 $t_{L2}$ 形成的柄-嗜喋结构和 *sib* 位点形成附加的柄-嗜喋结构；
（左下）从 $P_I$ 起始的转录物只具有 $t_{L2}$ 所形成的柄-嗜喋结构。

当生长条件迫使 λ 噬菌体进入溶原生长时，由于 CⅡ 蛋白的帮助，从 $P_I$ 起始转录整合酶基因。这时 RNA 聚合酶没有受到 N 蛋白的修饰，转录至终止子 $t_{L2}$ 处即告终止。由于 $t_{L2}$ 的转录物形成具有 11 个碱基对的柄-嗜喋结构，从而保护了整合酶 mRNA 的 3′ 末端免遭外切核酸酶的攻击（图 8-31）。由于 $P_I$ 位于 *xis* 基因内，因此从 $P_I$ 发起的转录只能表达整合酶基因，而没有切除酶的产生。

当原噬菌体受到外界因素的诱导而将要脱离细菌染色体时，这一过程的实现需要切除酶

*xis* 基因的表达，也需要整合酶 *int* 基因的表达。这两个基因的转录物都包含在从 $P_L$ 起始转录的 $mRNAL_2$ 中。这个 $L_2$ 与游离噬菌体所产生的 $L_2$ 不同，它不含 *sib* 位点(因为 λ 噬菌体整合的附着位点 *att* 位于 *int* 基因的编码序列与 *sib* 位点之间)，因而也就不存在返回调控。这样，左向转录物 $L_2$ 中的切除酶和整合酶都能够得到表达。这两种基因产物正是 λ 噬菌体从溶原状态转变为裂解生长的过程所必需(见第十一章)。

关于内切核酸酶参与 mRNA 降解的另一个例子是大肠杆菌编码 S21-引发酶-RNA 聚合酶 σ 亚基的操纵元(图 8-18)。这一多基因 mRNA 中，引发酶部分最不稳定(即降解最快)。据认为，这是首先由 RNA 内切酶在引发酶编码区和 σ 亚基编码区之间切断，然后由外切酶从 3′ 末端发生降解，但 S21 和 σ 亚基编码区的 3′ 末端均有发夹结构，可以抵抗外切酶的作用，因而表现出相当大的稳定性。再加上 S21 基因和引发酶基因之间具有衰减子结构，因而细胞内核糖体蛋白质 S21、引发酶、RNA 聚合酶 σ 亚基的分子数分别为 20 000、50 和 3 000。由此可见，在多基因 mRNA 中，各个基因的表达量不但互不相同，而且其差异也没有一个简单的模式(如 *lacZYA* 那样由 5′ 端至 3′ 端的基因产物依次下降，其比例为 1:0.5:0.2)。

综上所述，在细菌中参与 mRNA 降解的内切酶主要是 RNAase Ⅲ，它需要一定二级结构特征，而不是一个随机的内切酶。mRNA 或经过 RNAase Ⅲ 切断的 mRNA 片断都可以被 3′-外切核酸酶降解。就目前所知，细菌中存在着两种这样的3′-外切核酸酶，即 RNAase Ⅱ 和多聚核苷酸磷酸化酶。mRNA 中的终止子结构以及与终止子结构类似的发夹结构均能以不同的程度阻碍 3′-外切核酸酶对 mRNA 的降解。

### 四、蛋白质合成的自体调控

细胞控制蛋白质基因的表达，除了在转录水平上控制 mRNA 的产生之外，还可以通过各种花样控制 mRNA 的可翻译性。关于 mRNA 的二级结构对于翻译起始的影响，已在上面谈过。有些蛋白质还能直接控制自身 mRNA 的可翻译性。这类蛋白质，就目前所知，大多是核酸(RNA 或 DNA))结合蛋白或者与核酸分子相互作用作为其生理功能的蛋白质。也就是在这些蛋白质的 mRNA 上也有其结合位点。这里，我们只讲述大肠杆菌的蛋白质合成的释放因子 RF2 和核糖体蛋白质合成的自体调控。

1．*释放因子 RF2 合成的自体调控* 现在已经了解的对于 mRNA 可翻译性的控制，大多是控制翻译的起始，即控制核糖体与 mRNA 的结合。下面我们将要讲到的核糖体蛋白质合成的翻译控制就属于这种情况。然而，近年来人们又了解了其他的翻译控制手段。1985 年人们所发现的 RF2 对于自身 mRNA 翻译的控制就是利用它释放肽链的职能提前终止其 mRNA 的翻译。这到底是怎么回事呢?

原来，RF2 这个 340 个氨基酸的蛋白质，它们的氨基酸密码子并不处于同一阅读框中。其前 25 个氨基酸密码子处于 AUG 所在阅读框中，而其余的 315 个氨基酸的密码子处于 +1 阅读框中；在这两个编码区域之间多了一个 U，这个 U 与第 26 个氨基酸(Asp)的密码子GAC 的前两个核苷酸构成了 RF2 所能识别的终止密码子 UGA，而且是前一个编码区的同框终止密码子(图 8-32)。当细胞内 RF2 供应充足时，当核糖体 A 位到达第 25 个密码子后面的 UGA 时，RF2 就促成了肽链合成的终止。这是不难理解的。另一方面，如果细胞内缺乏 RF2，则核糖体以某种尚不知晓的机制向前滑动一个核苷酸，也就是发生了移框作用，继续合

成第 26 个氨基酸直到基因最后的终止密码子 UAG。显然，UAG 可由另一个释放因子 RF1 识别并促成 RF2 肽链的释放。现在还不知道，为什么这样一种精巧的自体调控机制只被RF2基因所利用，而 RF1 仍然是由常规的连续开放阅读框编码并以 UGA 作为终止密码子。RF2 基因表达的自体调控可能与保持 UGA 密码子的灵活性有关。

通过饱和突变试验，人们发现 RF2 mRNA 上的这个移框“窗口”至少得保留 15 个核苷酸，即 5′-AGG-GGG-UAU-CUU-UGA-3′，这样才能保证 30% 以上的移框率。实验证明，AGGGGG 这个类似于 S-D 序列的部分对于移框作用十分重要，任何一个碱基替代突变都使移框效应降低 5 倍以上。这种突变可为 16S rRNA3′ 末端的能继续维持配对状态的相应突变所补偿。这就表明，这一内部 S-D 序列仍然是与 16S rRNA3′ 末端的富含嘧啶区域(3′UCCUCC5′)发生相互作用。这两段序列的配对又是如何促成移框作用的，现在仍然不清楚。

移框窗口

5′...GUUCUUAGGGGGUAUCUUUGACUACGAC...3′

$NH_2$..ValLeuArgGlyTyrLeu AspTyrAsp...COOH

20 21 22 23 24 25 26 27 28

图 8-32 RF2 mRNA 移框窗口及其附近的结构。0 阅读框的终止密码子 UGA 能为 RF2 所识别而实现翻译的终止；在缺少 RF2 时，能通过移框作用连续翻译后面 315 个氨基酸，产生完整的 RF2 蛋白质分子

2. 核糖体蛋白质合成的自体调控　大肠杆菌的基因表达机构中约有 70 余种蛋白质。其中核糖体蛋白质最多，有 50 多种；还有若干种辅助蛋白质以及 RNA 聚合酶的亚基和辅因子。这些蛋白质合成的协调控制使细胞与生长条件相适应。大肠杆菌中有许多协调这些基因表达的机制，可能有一些机制至今还没有为人们所认识。

这类基因均为单拷贝，它们混合编组构成若干个操纵元，现将已知的六个操纵元的组成情况列于表 8-8。每个操纵元均以其中第一个已知功能的基因来命名。每个操纵元含有不同的基因。在大多数情况下，一个操纵元中不同基因的产物并不表现功能上的相关性，因为这些基因产物是在一个更大的整体机构中起作用的。但是 RNA 聚合酶的 $\beta$ 亚基和 $\beta'$ 亚基的基因(*rpoB* 和 *rpoC*)仍然排列在同一个操纵元中。在染色体上，前四种操纵元排列在一起(*str*-14 000bp-s10-*spc*-$\alpha$)，而后两个操纵元紧密相连。

绝大多数核糖体蛋白质(除了 L7/L12，它在每个核糖体中有 4 个分子)每个核糖体中只有一个分子，因此它们的表达必须与 rRNA 相协调。对于延伸因子 EF-Tu，每个细胞中 Tu 分子数大约是核糖体数的 10 倍。RNA 聚合酶的每种亚基数要比核糖体数目少一些。由于这些基因混合编组在不同的操纵元中，就必定存在着一种协调机制，让上述需要不同表达量的基因各得其所。

在表 8-8 的六个操纵元中，每个操纵元都有一个自己的调控蛋白质。这种调控蛋白质全部为核糖体蛋白质，而且都是在核糖体中直接与 rRNA 相结合的蛋白质。进一步分析发现，在这些 mRNA 上与调控蛋白质相结合的序列与 rRNA 上这一蛋白质结合的序列有很大的同源性，并有形状相似的二级结构。二者都能与起调控作用的核糖体蛋白质相结合，但是

表8-8 核糖体蛋白质、蛋白质合成因子、RNA聚合酶亚基等蛋白质的基因散布在几个操纵元中。每个操纵元的启动子均位于左端。带底线的是受调节蛋白质调控的基因,带点线的表示尚未肯定受调控的基因。

| 操纵元 | 基因及其蛋白质 | 调控蛋白质 |
|---|---|---|
| str | *rpsL–rpsG–fusA–tufA*<br>S12 S7 EF-G EF-Tu | S7 |
| spc | *rplN–rplX–rplE–rpsN–rpsH–rplF–rplR–rpsE–rplO–rpmD*<br>L14 L24 L5 S14 S8 L6 L18 S5 L15 L30 | S8 |
| S10 | *rpsJ–rplC–rplB–rplD–rplW–rplS–rplV–rpsC–rpsQ–rplP–rpmC*<br>S10 L3 L2 L4 L23 S19 L22 S3 S17 L16 L29 | L4 |
| α | *rpsM–rpsK–rpsD–rpoA–rplQ*<br>S13 S11 S4 α L17 | S4 |
| L11 | *rplK–rplA*<br>L11 L1 | L1 |
| rif | *rplJ–rplL–rpoB–rpoC*<br>L10 L7/L12 β β' | L10 |

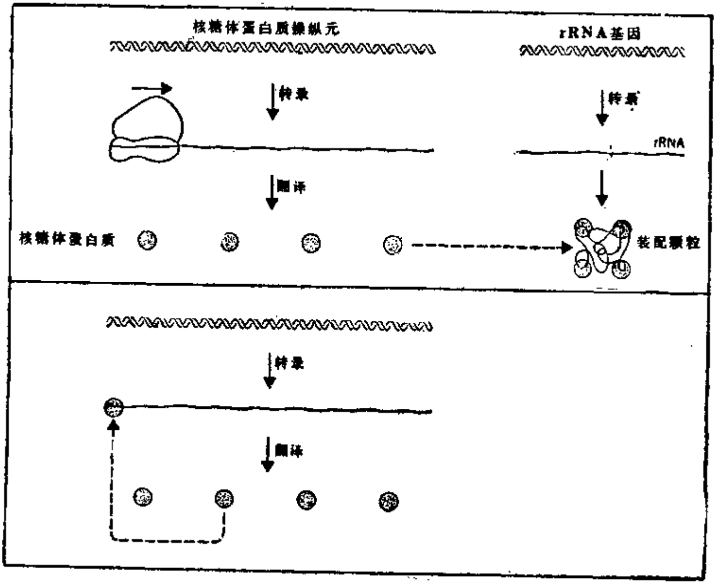



图8-33 大肠杆菌核糖体蛋白质翻译的自体调控
上图:有游离rRNA存在;下图:没有游离rRNA存在

rRNA 的结合能力大于各 mRNA。当细胞内核糖体蛋白质较少时，也就是说有游离的 rRNA 存在时，这些调控蛋白质优先与 rRNA 结合。当 rRNA 被饱和之后，多余的调控蛋白质就与 mRNA 上的结合位点结合。这个结合位点包含或靠近 S-D 序列，因此核糖体就不能结合，从而降低了有关基因产物的合成。现将大肠杆菌核糖体蛋白质合成的自体调控过程示于图8-33。

至于为什么在同一个操纵元中有的基因产物受到控制，有的却不受控制？这主要决定于各个基因是否有自己的 S-D 序列。如果有独立的 S-D 序列，则不受调控蛋白质的控制。这也就是表 8-8 的每个操纵元中受到控制的基因为什么总是连在一起的道理。以 L11 操纵元为例，调节蛋白质 L1 与操纵元中第一个基因 *rplk*(L11)的起始密码子上游的一段序列结合(图 8-34)，从而妨碍了核糖体与 S-D 序列结合。由于 L11 基因和 L1 基因之间只有 3 个核苷酸的间隔，L1 没有自己的核糖体结合位点，其翻译只能依靠从 L11 翻译过来的核糖体。因此，L1 蛋白质对于 L11 mRNA 翻译的抑制也同时抑制了 L1 自身的翻译。

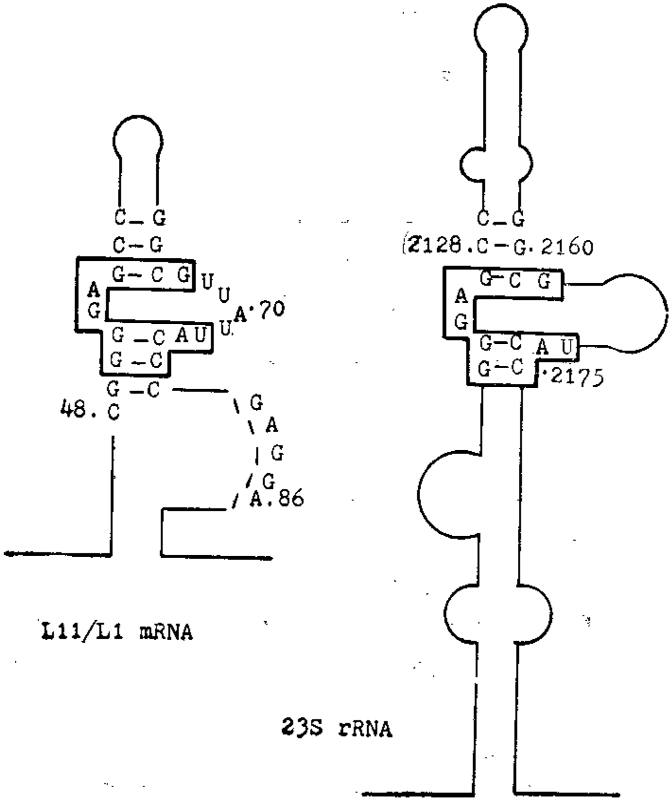



图 8-34 核糖体蛋白质 L1 在 23S rRNA 上的结合位点与 L11/L1mRNA 上的结合位点的比较。图中 GAGGA 为 L11/L1 mRNA 的 S-D序列。

这里，说的是 RNA(rRNA) 的合成水平调控了核糖体蛋白质基因的表达。上一节中我们讲过蛋白质合成水平通过衰减子系统影响到核苷酸合成酶的表达从而控制 RNA 的合成。由此可见在生物体内核酸的合成和蛋白质的合成是相互协调的，相互协调的信息是双向的，而不是单向的。下面我们将要谈到一种更为复杂的调控信息网络。

### 五、严谨反应

当细菌在不良的营养条件下生长时，由于缺乏足够的氨基酸，蛋白质合成受到抑制，并由此影响到细胞的许多生理生化活性，代谢水平下降，生长速度变慢。细菌对于不良的营养条件所产生的这一系列的反应叫做严谨反应(stringent response)。细菌为了渡过困难时期，在严谨反应中关闭许多生理活动。稳定 RNA(tRNA 和 rRNA)的合成速度下降了 10～20 倍，从而使 RNA 合成水平下降到正常状态下的 5～10%。部分种类的 mRNA 合成减少，但 mRNA 的总合成量减少约 3 倍。蛋白质降解速度增加。核苷酸、碳水化合物、脂类的合成均明显减少。严谨反应不仅调控了蛋白质合成，而且也调控基因转录、DNA 复制等许多细胞生理过程。

缺乏任何一种氨基酸或引起任何一种氨酰基 -tRNA 合成酶失活的突变都能导致严谨反应。这就说明严谨反应的触发器是位于核糖体 A 位的无负载的 tRNA。当这种无负载的 tRNA 进入 A 位以后，无法形成新的肽键，而 GTP 却在不断地消耗。这就是所谓空转反应(idling reaction)。细胞内出现空转反应时，就发出一种报警信号，这就是鸟苷 -5′- 二磷酸 -3′- 二磷酸(ppGpp)和鸟苷 -5′- 三磷酸 -3′- 二磷酸(pppGpp)。人们很早就发现，当大肠杆菌处于氨基酸饥饿时，在体内出现了两种异常的核苷酸，在薄层层析图谱上的迁移率与常见的核苷酸不同，人们称之为魔斑 I 和魔斑 II。后来才知道，魔斑 I 就是 ppGpp，魔斑 II 就是 pppGpp。

ppGpp 和 pppGpp 又是如何产生的呢？人们通过遗传学方法分离到不表现严谨反应的突变体，即所谓松弛型突变体(relaxed mutants)。各种松弛型突变位点分布在几个不同的基因中，其中大部分分布在 *relA* 基因中。*relA* 基因编码一种蛋白质，称为严谨因子。在正常情况下 *relA* 基因表达很少，大约 200 个以上的核糖体中才有一个核糖体结合有一个严谨因子。当氨基酸饥饿时，*relA* 基因的表达反而增加。松弛型突变除了发生在 *relA* 基因以外，还可以出现在核糖体 50S 亚基蛋白质 L11 的基因中以及 tRNA 基因的 TΨC 相应位置。这就说明，生成魔斑的反应不仅需要 *relA* 基因产物，而且还需要核糖体蛋白质 L11 和 tRNA 的 TψC 区。*relA*$^-$ 松弛型细胞提取液不能合成 ppGpp 和 pppGpp，而只要加入严谨因子就能合成。人工合成的四核苷酸 TΨCG 能够代替 tRNA 产生魔斑。而 L11 并不能代替核糖体，可能魔斑的合成需要核糖体的整体结构或涉及到 L11 以外的其他核糖体蛋白质。大肠杆菌的严谨反应中 ppGpp 和 pppGpp 的合成途径示于图 8-35。

ppGpp 和 pppGpp 是如何调控基因转录的，现在还不清楚。已知 ppGpp 能够与 RNA 聚合酶结合。然而为什么 tRNA 和 rRNA 的转录受到更大的抑制呢？这可能因为这两种稳定 RNA 基因的启动子与蛋白质基因不同。在蛋白质基因中，不同的基因所受到的抑制也不相同。这些现象目前尚无法解释。可能这两种异常的核苷酸并不直接与基因发生作用，很可能通过与某种因子结合而发生作用。总之，ppGpp 和 pppGpp 能通过某种方式控制细胞的许多生理过程，以使细胞能够适应有限的营养条件(特别是氨基酸的供应)。

$$pppG + ATP \xrightarrow{\text{严谨因子}} pppGpp + AMP$$

↓ 各种核糖体蛋白质，EF-Tu 和 EF-G

$$ppG + ATP \xrightarrow{\text{严谨因子}} ppGpp + AMP$$

$$+$$

$$Pi$$

图 8-35 大肠杆菌在严谨反应中魔斑生成途径。
实线表示主要途径，虚线表示次要途径。

前面我们已经讲过，当细胞的碳源供应不足时，细胞内产生 cAMP。当细胞受到氨基酸饥饿时，细胞内则产生 ppGpp 和 pppGpp。这些非同寻常的代谢产物能够影响某些蛋白质和酶的活性或性质，从而设法解决细胞所碰到的问题。这些非同寻常的代谢产物，我们称之为报警素(alarmones)。细胞内可能存在着许多报警素。近来人们在原核细胞和真核核细胞中都已发现双腺苷四磷酸(ApppppA)。当原核细胞内 tRNA 短缺时，由氨酰基 -tRNA 合成酶催化产生 ApppppA。真核细胞中当复制叉停顿时也会产生ApppppA。ApppppA 能够刺激真核生物 DNA 聚合酶 $\alpha$ 的活性以促进 DNA 的合成。

## 第八节 DNA 序列重排对基因转录的调控[68]

在少数细菌中，某些基因的表达与 DNA 序列的重排有关。这方面研究得较多的是鼠伤寒沙门氏菌的鞭毛基因的相变过程。

沙门氏菌的鞭毛蛋白由两个不同的基因 *H*1 和 *H*2 所编码。一个细菌中若是由 *H*1 基因表达产生 H1 型鞭毛蛋白，这时细菌处于 I 相(phase I)；若是由 *H*2 基因表达产生 *H*2 型鞭毛蛋白，则细菌处于 II 相(phase II)。处于 I 相或 II 相细菌在进行细胞分裂时，以 1/1000 的频率产生另一相的后代，这个过程就叫做相变(phase variation)。

*H*1 基因和 *H*2 基因位于染色体上的不同区域。*H*2 基因与另一个编码 *H*1 基因阻遏蛋白的基因 *rh*1 紧密连锁，这两个基因协调表达。当 *H*2 基因表达时，*rh*1 也同时表达，*rh*1 基因产物就阻碍了 *H*1 基因的表达。当 *H*2 基因不表达时，*rh*1 也不表达，这时 *H*1 基因就表达。因此，沙门氏菌的相决定于 *H*2-r*h*1 转录单元是否表达(图 8-36)。

*H*2-*rh*1 转录单元是否表达是受到其上游一段 DNA 序列的排列方向所控制。这段DNA序列为 995bp，其两端各含有一段 14bp 长的倒转重复序列(IRL 和 IRR)。*H*2 基因的起始密码子与这段 995bp 的 DNA 序列相隔 16bp，但启动子位于 995bp 序列的末端。在这段序列中，有个基因叫做 *hin* 基因，其蛋白质产物可以催化这段 995bp DNA 的倒位(图 8-37)。当启动子序列朝向 *H*2 基因时，则*H*2-*rh*1 转录单元表达，而基因*H*1 不能表达，细菌处于 II 相(图 8-37 上)。当 995bp 序列倒位后，启动子移动到另一头，而且背离 *H*2 基因，因此 *H*2-*rh*1 转录单元不能表达，这时 *H*1 基因表达，细菌就处于 I 相(图 8-37 下)。关于倒位的机制将在第十一章中阐述。

在人工培养基中生长的沙门氏菌，其 I 相和 II 相并无什么生理上的差异。相变的意义在于逃脱寄主抗体的进攻。例如细菌侵染寄主时处于 I 相，则产生 H1 型鞭毛蛋白，寄主就

可能针对 H1 型蛋白而制造抗体，而细菌利用相变的手段产生 1/1000 的 Ⅱ 相细菌后代，又可以在寄主体内生存和繁衍。

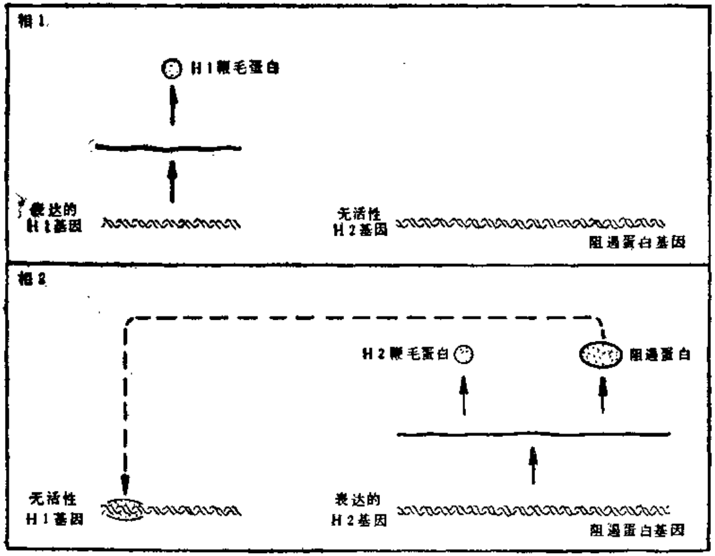



图 8-36 沙门氏菌相变取决于 *H*2-*rh*1 转录单元的表达与否

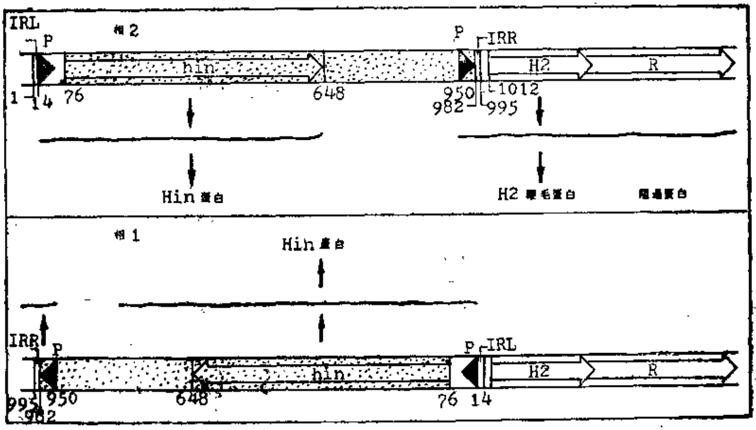



图 8-37 995bp 的可倒位序列的方向控制了 $H_2$ 鞭毛蛋白和 $H_1$ 阻遏蛋白的表达

## 提　要

原核生物是单细胞生物，与外界环境密切相关。它们必须不断地调整各种不同基因的表达以适应营养条件和应付各种不利的物理化学因素。基因表达的调控主要发生在转录水平上，在其它水平上也常出现各种不同的调控方式。

根据操纵元对调节蛋白质的响应情况，可以区别负控制系统和正控制系统。当加入调节蛋白质后基因表达活性被关闭的控制系统就叫做负控制系统，其调节蛋白质称为阻遏蛋白；当加入调节蛋白质后基因表达活性被开启的控制系统则叫做正控制系统，其调节蛋白质称为无辅基诱导蛋白。根据操纵元对于能调节它们表达的小分子的应答反应的性质，可以区分可诱导的操纵元和可阻遏的操纵元。小分子物质可以开启可诱导操纵元的基因转录活性，相反小分子物质可以关闭可阻遏操纵元的基因转录活性。

Jacob 和 Monod 根据对 *lac* Z, Y, A 基因突变体的研究，于 1961 年提出了操纵元模型。其要点是：一个或几个结构基因与一个调节基因和一个操纵位点组成一个操纵单元。这个单元就称其为操纵元，这个操纵位点就称为操作子。调节基因产生的阻遏蛋白与操纵位点结合从而阻碍了结构基因转录成为 mRNA，而诱导物又可以与阻遏蛋白相结合从而阻止阻遏蛋白与操作子的结合。后来人们发现的参与转录起始的启动子也是操纵元的一个成分。

乳糖操纵元模型提出以后，人们相继了解了许多其它的操纵元，这些操纵元都各有其特点。例如半乳糖操纵元具有双启动子结构；阿拉伯糖操纵元的调节蛋白质具有正控制和负控制的双重功能；色氨酸操纵元是一个可阻遏的操纵元。然而，后来人们发现这些操纵元的某些"特点"带有一定的普遍性，例如乳糖操纵元也有双启动子结构，乳糖操纵元与半乳糖操纵元和阿拉伯糖操纵元一样，也具有一个以上的阻遏蛋白结合位点。

原核生物在生长发育的各个阶段，基因的表达是按照一定的时间顺序而展现的。这种时序调控，大多是通过一种或几种蛋白质与 RNA 聚合酶相互作用而实现的，如枯草杆菌及其噬菌体中 $\sigma$ 亚基的更迭，大肠杆菌热震惊基因的表达，$\lambda$ 噬菌体的抗终止作用，$\lambda$ 噬菌体启动子 $P_E$ 和 $P_I$ 的识别等。有时，整个 RNA 聚合酶被更换，如 T7 噬菌体的第 Ⅱ 组和第 Ⅲ 组基因的表达。

原核生物没有核膜，转录过程和翻译过程得以耦联。这是原核生物中衰减子调控的基础。衰减子调控的实质是以翻译手段调控基因的转录，也就是通过某种机制影响核糖体与 RNA 聚合酶在 mRNA 上的相对位置从而控制终止子的二级结构是否形成或形成的二级结构所持续的时间。不少氨基酸或核苷酸合成操纵元均存在着衰减子结构。

在原核生物中，翻译水平上调控也是十分重要的调控。反义 RNA 通过与 mRNA 的S-D 序列以及起始密码子等部分互补，从而封闭了 mRNA 的翻译活性。与反义 RNA 的作用相似，mRNA 分子本身也可以形成二级结构从而影响翻译的进行。mRNA 的寿命也显著地影响到蛋白质的合成，$\lambda$ 噬菌体整合酶mRNA的返回调控就是很好的例子。除了 RNA 二级结构可以控制 mRNA 的可翻译性之外，某些基因的产物能够与其 mRNA 结合，从而控制这种蛋白质的继续合成，如释放因子 RF2 和核糖体蛋白质的自体调控。在翻译过程中，当无氨基酸负载的 tRNA 进入 A 位后出现的空转反应促进了严谨因子的表达，从而产生魔斑。魔斑又以尚不知晓的机制调整细胞的生理活动。

原核生物中还有以 DNA 重排的方式控制基因转录的，鼠伤寒沙门氏菌的鞭毛蛋白基因

的相变过程就是一例。

## 参 考 文 献


1. Jacob, F. and Monod, J. (1961)"Genetic Regulatory Mechanisms in the Synthesis of Proteins." *J. Mol. Biol.* 3: 318～356.
2. Gilbert, W. and Müller-Hill, B. (1966) "Isolation of the Lac Repressor." *Proc. Nat. Acad. Sci.* 56: 1891～1898.
3. Ptashne, M. (1967) "Specific Binding of the λ Phage Repressor to λ DNA." *Nature* 214:232～234.
4. DeCrombrugghe, B. et al. (1984) "Cyclic AMP Receptor Protein: Role in Transcription Activation." *Science* 224: 831～838.
5. Raibaud, O. and Schwartz, M. (1984) "Positive Control of Transcription initiation in Bacteria." *Ann. Rev. Genetics* 18 :173～206
6. Miller, J. and Reznikoff, W. eds (1980) *The Operon*, Cold Spring Harbor Laboratory, New York.
7. Straney, S. B. and Crothers, D. M. (1987) "Lac Repressor Is a Transient Gene-activating Protein." *cell* 51: 699～707.
8. Gottesman, S. (1984) "Bacterial Regulation: Global Regulatory Networks." *Ann. Rev. Genetics* 18 :415～441.
9. Majumdar, A. and Adhya, S. (1984) "Demonstration of Two Operator Elements in *gal*: in Vitro Repressor Binding Studies" *Proc. Nat. Acad. Sci* 81: 6100～6104.
10. Dunn, T. M. et al.(1984)"An Operator at -280 Base Pairs that Is Required for Repression of *araBAD* Operon Promoter: Addition of DNA Helical Turns Between the Operator and Promoter Cyclically Hinders Repression." *Proc. Nat. Acad. Sci.* 81: 5017～5020.
11. Hendrickson, W. and Schleif, R. (1985) "A Dimer of AraC Protein Contacts Three Adjacent Major Groove Regions of the *araI* DNA Site." *Proc. Nat. Acad. Sci.* 82: 3129～3133.
12. Platt, T. (1987)"Regulation of Gene Expression in the Tryptophan Operon of *Escherichia coli*."in *The Operon*, ed. Miller, J. H. and Reznikoff, W. S. Cold Spring Harbor Laboratory. New York, pp. 263～302.
13. Yanofsky, C. et al. (1981) "The Complete Nucleotide sequence of the Tryptophan Operon of *E. Coli*." *Nucleic Acid Res.*, 9: 6647～6668
14. Helmann, J. D. and Chamberlin, M. J. (1988) "Structure and Function of Bacterial Sigma Factors." *Ann. Rev. Biochem.* 57: 839～872
15. Losick, R. and Pero, J. (1981) "Cascades of Sigma Factors." *Cell* 25: 582～584.
16. Fox, T. D. et al. (1976) "Regulatory Gene 28 of Bacteriophage SPOl Codes for a Phage-induced Subunit of RNA Polymerase." *J. Mol. Biol.* 101: 427～433.
17. Costanzo, M. et al. (1984) "Bacteriophage SPOl Genes 33 and 34: Location and Primary Structure of Genes Encoding Regulatory subunits of *Bacillus subtilis* RNA Polymerase." *J. Mol. Biol.* 180:533～547.
18. Johnson, W. et al. (1983) "Two RNA Polymerase Sigma Fators from *Bacillus subtilis* Discriminate Between Overlapping Promoters for a Developmentally Regulated Gene." *Nature* 302: 800～804.
19. Neidhardt, F. C. et al. (1984) "Genetics and Regulation of Heat Shock Proteins." *Ann. Rev. Genetics* 18: 295～329.
20. Grossman, A. et al. (1984) "The *htpR* Gene Product of *E. coli* Is a Sigma Factor for Heat-shock Promoters." *Cell* 38: 383～390.
21. Cowing, D. W. et al.(1985) "Consensus Seqence for *Escherichia coli* Heat Shock Gene Promoters." *Proc. Nat. Acad. Sci.* 82 2679～2683.
22. Yura, T. et al. (1984) "Heat Shock Regulatory Gene (*htpR*) of *Escherichia coli* Is Required for Growth at High Temperature but Is Dispensable at Low Temperature." *Proc. Nat. Acad. Sci.* 81: 6803～6807.
23. Hirschman, J. et al. (1985) "Products of Nitrogen Regulatory Genes *ntrA* and *ntrC* of Enteric Bacteria Activate *glnA* Transcription in Vitro: Evidence that the *ntrA* Product Is a σ Factor." *Proc. Nat. Acad. Sci.* 82: 7525～7529.
24. Finley, D. et al. (1987) "The Yeast Polyubiquitin Gene Is Essential for Resistance to High Temperature, Starvation, and Other Stresses." *Cell.* 48: 1035～1046.
25. Elliot, T. and Geiduschek, E. P. (1984) "Defining a Bacteriophage T4 Late Promoter: Absence of a "-35" Region." *Cell* 36: 211～219.

26. Kassavetis, G. and Geiduschek, E. (1984) "Defining a Bacteriophage T4 Late Promotor: The T4 Gene 55 Protein Suffices for Directing Late Promoter Recognition." *Proc. Nat. Acad. Sci.* 81: 5101〜5105.
27. Dunn, J. J. and Studier, F. W. (1983) "Complete Nucleotide Sequence of Bacteriophage T7 DNA and the Locations of T7 Genetic Elements." *J. Mol. Biol.* 166: 477〜535.
28. Chamberlin, M. et al, (1970) "New RNA Polymerase from *E. coli* Infected with Bacteriophage T7." *Nature* 228: 227〜231.
29. Mark, K. and Studier, F. W. (1981) "Purification of the Gene 0.3 Protein of Bacteriophage T7, an Inhibiter of the DNA Restriction System of *Escherichia coli.*" *J. Biol. Chem.* 256: 2573〜2578.
30. Bailey, J. N. et al. (1983)"Relationship Between Promoter Structure and Template Specificities Exhibited by the Bacteriophage T3 and T7 RNA Polymerases." *Proc. Nat. Acad. Sci.* 80: 2814〜2818.
31. Studier, F. W. and Dunn J. J. (1982) "Organization and Expression of Bacteriophage T7 DNA." *Cold Spring Harbor Symp Quant. Biol.* 47: 999〜1007.
32. Hendrix, R. W. et al. eds. (1983) *Lambda* II, Cold Spring Harbor Laboratory, New York.
33. Johnson, A. D. et al. (1987)"λ Repressor and Cro-components of an Efficient Molecular Switch."*Nature* 294: 217〜223.
34. Meyer, B. J. et al. (1980) "Gene Regulation at the Right Operator ($O_R$) of Bacteriophage λ. II. $O_R1$, $O_R2$, and $O_R3$: Their Roles in Mediating the Effects of Repressor and Cro." *J. Mol. Biol.* 139: 163〜194.
35. Eisen, H. et al. (1970) "Regulation of Repressor Expression in λ." *Proc. Nat. Acad. Sci.* 66: 855〜862.
36. Herskowitz, I. and Hagen, D. (1980) "The Lysis-lysogeny Decision of Phage λ: Explicit Programming and Responsiveness." *Ann. Rev. Genetics* 14: 399〜445.
37. Shimatake, H. and Rosenberg, M. (1981) "Purified λ Regulatory Protein CII Positively Activates Promoters for Lysogenic Development." *Nature* 292: 128〜132.
38. Hoyt, M. A. et al.(1982) "Control of Phage λ Development by Stability and Synthesis of CII protein: Role of the Viral CIII and Host *hflA*, *himA* and *himD* Genes." *Cell* 31: 565〜573.
39. Ilan, J. ed. (1987)*Trnaslational Regulation of Gene Expression*, Plenum Press, New York.
40. Yanofsky, C. (1988) "Transcription Attenuation." *J. Biol. Chem.* 263: 609〜612.
41. Yanofsky, C. and Kolter, R. (1982) "Attenuation in Amino Acid Biosynthetic Operons." *Ann. Rev. Genetics* 16: 113〜134.
42. Bauer, C. et al. (1983) "Attenuation in Bacterial Operons."in *Gene Function In Prokaryotes*, ed Beckwith, J. et al. Cold Spring Harbor Laboratory, New York, pp.65〜89.
43. Jensen, K. F. et al. (1986) "Attenuation at Nucleotide Biosynthetic Genes and Amino Acid Biosythetic Opersons of *Escherichia coli.*" *Trends Biochem. Sci.* 11: 362〜365.
44. Green, P. J. et al. (1988) "The Role of Antisense RNA in Gene Regulation." *Ann. Rev. Biochem.* 55: 569〜597.
45. Laporte, D. C. (1984) "Antisense RNA: a New Mechanism for the Control of Gene Expression." *Trends Biochem. Sci.* 9: 463.
46. Simons, R. and Kleckner, N. (1983) "Translational Control of IS10 Transposition." *Cell.* 34: 683〜691.
47. Mizuno, T. et al. (1984) "A Unique Mechanism Regulating Gene Expression: Translational Inhibition by a Complementary RNA Transcript." *Proc. Nat. Acad. Sci.* 81: 1966〜1970.
48. Rebagliati, M. R. and Melton, D. A. (1987) "Antisense RNA Injections in Fertilized Frog Eggs Reveal an RNA Duplex Unwinding Activity." *Cell* 48: 599〜605.
49. Hoopes, B. C. and McClure, W. R. (1985) "A CII-dependent Promoter Is Located within the Q Gene of Bacteriophage λ." *Proc. Nat. Acad. Sci.* 82: 3134〜3138
50. Gussin, G. N. et al. (1966) "Protein Synthesis Directed by RNA Phage Messengers." *Cold Spring Harbor Symp. Quant. Biol.* 31: 257〜272.
51. Fiers, W. et al. (1976) "Complete Nucleotide Sequence of Bacteriophage MS2 RNA: Primary and Secondary Structure of the Replicase Gene." *Nature* 260: 500〜507.
52. Gold, L. (1988) "Posttransciptional Regulatory Mechanisms in *Escherichia coli.*" *Ann. Rev. Biochem.* 57: 199〜233.
53. Brawermen, G. (1987) "Determinants of Messenger RNA Stability." *Cell* 48: 5〜6.
54. Belasco, J. G. et al. (1986) "The stability of *E. coli* Gene Transcripts Is Dependent on Determinants Localized to Specific mRNA Segments." *Cell* 46: 245〜251.
55. Gottesman, M. et al. (1982) "Retroregulation: Control of Gene Expression from Sites Distal to the Gene." *Cell*

29:727～728.

56. Roberson, H. D. (1982) "*Escherichia coli* Ribonuclease Ⅲ Cleavage Sites." *Cell* 30: 669～672.
57. Craigen, W. J. et al. (1985) "Bacterial Peptide Chain Release Factors, Conserved Primary Structure and Possible Frameshift Regulation of Release Factor 2 " *Proc. Nat. Acad. Sci.* 82 3616～3620.
58. Weiss, R. B et al. (1988) "Reading Frame Switch Caused by Base-pair Formation Between the 3′ End of 16S rRNA and the mRNA During Elongation of Protein Synthesis in *Escherichia coli*." *The EMBO* J. 7:1503～1507.
59. Nomura, M. et al. (1980) "Feedback Regulation of Ribosomal Protein Gene Expression in *Escherichia coli*: Structural Homology of Ribosomal RNA and Ribosomal Protein mRNA." *Proc. Nat. Acad. Sci.* 77: 7084～7088
60. Nomura, M. et al. (1984) "Regulation of the Synthesis of Ribosomes and Ribosomal Components." *Ann. Rev. Biochem.* 53: 75～117.
61. Baughman, G. and Nomura, M. (1984) "Translational Regulation of the L11 Ribosomal protein Operon of *Escherichia coli*: Analysis of the mRNA Target Site Using Oligonucleotide-directed Mutagenesis." *Proc. Nat. Acad. Sci.* 81 :5389～5393.
62. Starzyk, R. (1987) "A Target Site for Translational Regulation." *Trends Biochem, Sci.* 12: 415～418.
63. Gallant, J. A. (1979) "Stringent Control in *E. coli*." *Ann. Rev. Genetics* 13: 393～415.
64. Lamond, A. I. and Travers, A. A. (1985) "Stringent Control of Bacterial Transcription." *Cell* 41:6～8.
65. Stephens, J. et al. (1975) "Guanosine 5′diphospate 3′- diphosphate (ppGpp): Positive Effector for Histidine Operon-transcription and General Signal for Aminoacid Deficiency." *Proc. Nat. Acad. Sci* 72: 4389～4393.
66. Sarmientos, P. and Cashel, M. (1983) "Carbon Starvation and Growth Rate-dependent Regulation of the *Escherichia coli* Ribosomal RNA Promoters: Differential Control of Dual Promoters". *Proc. Nat. Acad. Sci.* 80. 7010～7013.
67. Lamond, A. I. and Travers, A. A. (1984) "Genetically Separable Functional Elements Mediate the Optimal Expression and Stringent Regulation of a Bacterial tRNA Gene." *Cell* 40: 319～326.
68. Simon, M. et al. (1980) "Phase Variation in the Evolution of a Controlling element." *Science* 209: 1370～1374.

# 第九章　真核生物基因表达的调控

## 第一节　概　述[1~6]

### 一、真核生物基因调控的特点

真核生物基因表达的调控系统比起原核生物有很大不同。首先，真核生物的基因数目比原核生物的多，大多数基因都含有内元。基因组中除了组成基因的 DNA 序列外还有各种非编码的功能尚不清楚的重复序列和单一序列，这些序列对基因表达可能带来影响。真核生物 DNA 与组蛋白形成核小体结构，同时还与许多非组蛋白结合。这些组蛋白、非组蛋白、核小体的组织状态及核小体的超螺旋结构等也都影响到基因活性。

真核生物中，转录和翻译分别在细胞核和细胞质中进行。这排除了类似于原核生物衰減作用存在的可能性。不过绝大多数真核生物 mRNA 的形成都要经过拼接加工，细胞可以通过决定mRNA 前体的加工与否以调控基因表达。mRNA 形成后，要输进细胞质才能翻译产生蛋白质。真核生物的 mRNA 寿命较长，这又为翻译水平的调控提供了较大的余地。所以，和原核生物主要在转录水平上进行调控的情况相比，真核生物基因表达的调控活动范围更大，包括 DNA 水平(基因拷贝数和重排)、转录水平、转录后水平(RNA 加工和运输)、翻译水平等许多层次的调控。然而，真核生物和原核生物一样，转录水平的调控是最主要的调控。

绝大多数真核生物都是多细胞的复杂有机体。由一个受精卵开始，在个体发育过程中逐步分化产生各种组织和细胞类型。分化是不同基因表达的结果。在正常情况下，一个个体的各种细胞类群总是按照一定的"计划"不断进行严格的调控，关闭某些基因，开启另一些基因，使个体发育得以顺利进行；而到了一定时期，在大多数细胞中分裂和生长停止，它们只需要维持自己。所以，不同细胞的基因表达依类型不同和所处发育阶段的不同而异，这和原核生物同一群体细胞内基因表达基本一致(若有差异亦多属偶然)的情况不同。

真核生物对外界环境条件变化的反应也和原核生物十分不同。同一群体的原核生物细胞处在同样的环境条件作用下，对环境条件变化的反映基本一致。而多细胞真核生物的不同细胞在同样环境条件下可能有很不相同的反应，只有一部分或很少一部分细胞的基因表达直接受到某一环境条件变化的影响和控制，其它细胞的基因表达或是间接地受到影响，或是基本上不受影响。真核生物的这一特点保证它的许多主要器官或组织在千变万化的环境下维持正常的功能。

应该指出的是，过去对真核生物基因表达的研究远远不如对原核生物的研究。其中的原因主要有：真核生物基因组成和形态结构复杂，难以直接鉴定基因产物和基因控制的生化过程；难以直接选择出影响调节基因的突变体；难以直接通过改变外界环境条件来研究分析基因表达的变化；难以直接进行基因操作等等。然而这些困难已经通过重组 DNA 技术、组织培养技术，定向诱变技术等新技术的广泛应用而逐步得到克服。

### 二、活跃表达基因的数目

为研究真核生物的基因表达，有必要对一个特定细胞类群中正常表达的基因数有一个粗

略的估计。

用 RNA 饱和杂交试验(RNA saturation hybridization) 可以求出在非重复序列 DNA 中能与 mRNA 杂交部分的百分率。将少量非重复序列 DNA 变性后与过量 mRNA 混合,使每个与 mRNA 互补的 DNA 序列都能有机会和 mRNA 形成双链杂合体。这种杂交又叫 RNA-驱动的杂交 (RNA-driven hybridization)。杂交程度为所用 RNA 浓度和反应时间的乘积(Rot)所制约。Rot越大。杂交就越彻底,直到趋于饱和。图 9-1 为海胆 *S. purpuratus* 非重复 DNA 和原肠胚多核糖体上 mRNA 杂交的情况。当 Rot 值超过 300 后,杂交趋于饱和。这时大

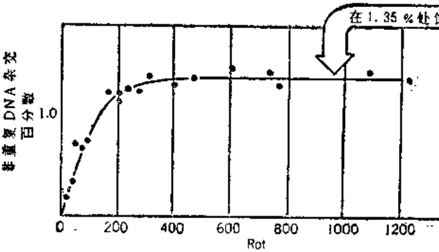



图 9-1 海胆 *S. purpuratus* 非重复 DNA 和过量原肠胚mRNA 的饱和杂交结果

约有 1.35% 的非重复 DNA 和 RNA 形成杂合体。由于 DNA 只有一条链和其 RNA 产物互补。所以上例 mRNA 群体代表 2.7% 的非重复 DNA 序列。海胆的 DNA 总量为 $8.1\times10^8$bp,其中非重复 DNA 占 75%,mRNA 群体所代表的 DNA 序列为 $0.027\times0.75\times8.1\times10^8=1.7\times10^7$ bp。已知该 mRNA 群体的平均长度为 2 000 碱基,则约有 $1.7\times10^7/2000=$ 8 500个基因在原肠胚细胞中处于活跃表达状态。

用上述方法测得,酵母中约有 4 000 个基因表达,高等真核生物的体细胞中约有 10 000~15 000 基因表达。由于 DNA 中存在着序列相同或相似的重复基因家族,上述饱和试验中形成的 DNA-RNA 杂合体可能包括了没有转录的序列,使估计的表达基因数目偏高。为了鉴定这一点,可以把 DNA-RNA 杂合体中的 DNA 提取出来用于 DNA 杂交试验中,看看这些 DNA 的 $Cot_{1/2}$值是否和非重复 DNA 的 $Cot_{1/2}$ 吻合。这种鉴定方法叫做回放试验(Play back experiment)。

### 三、基因表达的不同水平:丰富 mRNA 和稀少 mRNA

上面介绍了估计处于活跃表达状态基因数目的方法。那么,这些基因之间在表达水平上是否存在差异呢?这可以通过 RNA 驱动的动力学杂交试验分析 (图 9-2)。首先将 mRNA 群体提取出来并作为模板在 RNA 逆转录酶的作用下合成与 mRNA 互补的 cDNA。由于 cDNA 直接由 mRNA合成而来,某种 mRNA 在群体中的频率越高,由它产生的cDNA 在群体中的频率也就越高,在 cDNA 和过量 mRNA 的杂交中就越容易形成杂合双链。这和 DNA 杂交中高度重复序列容易复性的原理一样。一般把某一 cDNA 组分半数的分子与 mRNA 形成双链杂合体所需要的 Rot 值叫 $Rot_{1/2}$。通过比较待测 RNA 的 $Rot_{1/2}$ 值和对照(标记)RNA 的 $Rot_{1/2}$ 值,可以得出这个 RNA 组分的复杂长度。一般常用珠蛋白 cDNA-mRNA 的杂交作为对照。

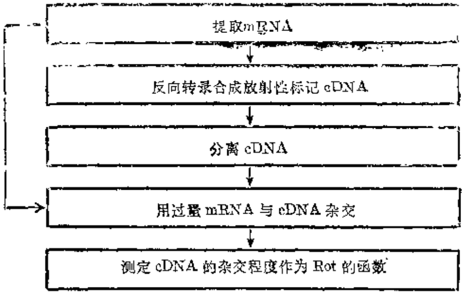



图 9-2 过量mRNA和cDNA杂交测定mRNA复杂性的基本步骤

$$\frac{\text{待测RNA复杂长度}}{\text{珠蛋白mRNA复杂长度}}=\frac{\text{待测RNA的}Rot_{1/2}}{\text{珠蛋白mRNA的}Rot_{1/2}}$$

这里的计算原理和方法都十分类似第二章中用比较 $Cot_{1/2}$ 对DNA复杂长度的计算。图9-3为鸡输卵管mRNA及其cDNA的杂交情况。在本试验中，所用的对照为提纯的鸡卵清

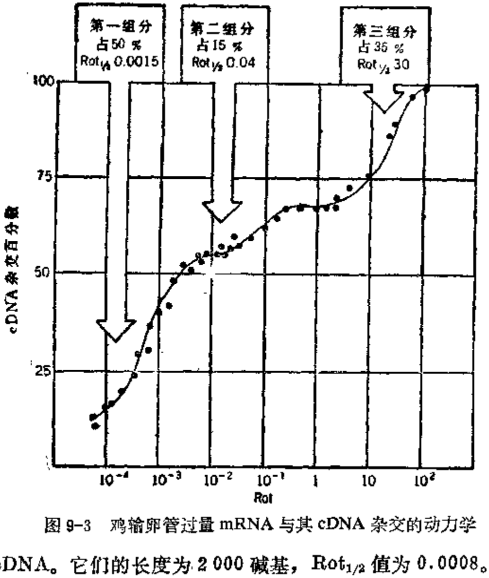



图 9-3 鸡输卵管过量mRNA与其cDNA杂交的动力学

蛋白mRNA和cDNA。它们的长度为2 000碱基，$Rot_{1/2}$ 值为0.0008。从图 19-3 可见，鸡输卵管的mRNA包括了三类不同的组分，最先杂交的部分 $Rot_{1/2}$ 值约为0.0015。占了全部mRNA(cDNA)的大约一半。如果把这个组分提纯单独进行杂交试验，所观察到的 $Rot_{1/2}$ 应该是 $0.5\times0.0015=0.00075$。这和作为对照的纯卵清蛋白mRNA的 $Rot_{1/2}$ 值0.0008非常相似。事实上这个组分正是卵清蛋白mRNA。这说明在鸡输卵管卵清蛋白mRNA大约占了全部mRNA的一半。

接着杂交的 RNA 组分约占了全部 RNA 的 15%，$Rot_{1/2}$ 值为 0.04。若经提纯，这个组分的实际 $Rot_{1/2}$ 应为 $0.15\times0.04=0.006$，其复杂长度为 $2\,000\times0.006/0.0008=15\,000$ 碱基。大概相应于 7～8 种平均长度为 2 000 碱基的 mRNA。

最后复性的组分占全部 RNA 的 35%。$Rot_{1/2}$ 值为 30。若经提纯，其 $Rot_{1/2}$ 值应为 $0.35\times30=10.5$。与之相应的复杂长度为 $2\,000\times10.5/0.0008=26\,000\,000$ 碱基。这相应于 13 000 种平均长度为 2 000 碱基的 mRNA。

从上面分析可知，在鸡输卵管组织的 mRNA 中，大约有 50% 为卵清蛋白 mRNA；另外的 7～8 种 mRNA 占了约 15%；而剩下的 35% 包括了大约 13 000 种 mRNA。由此可见，同一组织中不同种类的 mRNA 在拷贝数上差异一定很大。如果知道了一个细胞的 mRNA 总含量，可以通过计算求出某一组分 mRNA 的平均数目或丰富度 (abundance)：

$$丰富度=\frac{每细胞中\ mRNA\ 含量(克)\times待测\ mRNA\ 组分的百分比\times6\times10^{23}}{待测组分的复杂长度(道尔顿)}$$

仍以图 9-3 为例，鸡输卵管组织中每细胞含有 mRNA 0.275pg。对于第一组分 mRNA (卵清蛋白 mRNA)来说，每个细胞中约有 100 000 拷贝；而对于第二组分的 7～8 种 mRNA 来说，每一种在每细胞中约有 4 000 拷贝；而对于第三组分的 13 000 种 mRNA 来说，平均每一种大约只有 5 个拷贝。

绝大多数真核生物特定组织中 mRNA 的存在状况都和上述鸡输卵管的例子相似(虽然不是那么极端)，即少数(不到 100 种) mRNA 可达到数千甚至数万拷贝。它们往往占了全部 mRNA 的一半甚至更多。这类 mRNA 叫做丰富 mRNA (abundant mRNA) 或优势 mRNA (superprevalent mRNA)。在上例中卵清蛋白 mRNA 和第二组分 mRNA 是丰富 mRNA。另一方面，许多细胞中约一半的mRNA包括了数千甚至上万种，每种mRNA的拷贝数大多在 10 以下，它们叫做稀少 mRNA (scarce mRNA) 或复杂 mRNA (complex mRNA)。在RNA 饱和杂交试验中，正是这类稀少 mRNA 驱动了 DNA 的杂交。

通过动力学杂交试验计算的 mRNA 种类数目可能比实际的小些。因为可能某些含量过低的 mRNA 未被测出。一般把用这种办法得到的数目作为实际 mRNA 种类数目的下限，而把通过前面(如图 9-1)所述的饱和试验所估计的 mRNA 数目作为上限。这样，鸡输卵管中约有 13 000～15 000 种 mRNA。

### 四、持家基因和奢侈基因

高等真核生物细胞中活跃表达的基因一般有10 000～20 000个，例如通过饱和杂交试验测得的活跃表达基因数目在鸡输卵管中约为15 000，在鸡肝中约为17 000。这些基因在不同组织中有多少是相同的(重叠的)，多少是组织特异性的呢？

鸡输卵管中卵清蛋白 mRNA 占了全部的 50%，而在肝细胞中这种 mRNA 却根本不存在。所以，若从绝对质量来说、这两种组织的 mRNA 至少有一半不同。但现在的问题是要求出 mRNA 种类在这两种组织的不同，这里除了包括为数不多的丰富 mRNA 外，还包括一部分稀少 mRNA。

现在尚不能对组织中 mRNA 的重叠情况进行直接测定，但可以用间接的方法估计。常常采用的一种方法是加性饱和杂交试验。用鸡非重复 DNA 序列分别与来自肝和输卵管的 mRNA 杂交。能够参与杂交的部分分别占非重复 DNA 的 2.05% 和 1.80%。假若这两个

mRNA 样本所代表的 DNA 序列完全不同，则它们混合应该能够和 2.05% + 1.80% = 3.85% 的非重复 DNA 序列杂交。然而实际上这个数字是 2.4%。这说明在肝细胞中表达而在输卵管中不表达的基因数目约为 17 000 × (2.4 − 1.8)/2.05 ≈ 5 000，在输卵管中表达而在肝细胞中不表达的基因数目为 15 000 × (2.4 − 2.05)/1.80 ≈ 3 000，其余 12 000 种基因在两种组织中都表达。

测定不同组织活性基因重叠情况的另一种方法是用来自一种组织的过量 mRNA 与非重复 DNA 杂交。凡能与 mRNA 杂交的 DNA 序列叫做 mDNA。然后把参与杂交的 mDNA 组分和未参与杂交的无效 DNA(null DNA) 组分分开。分别用它们与来自第二种组织的过量 mRNA 杂交。mDNA 中和第二种 mRNA 杂交的百分比代表了在两种组织中都表达的基因(重叠基因)的百分比；而无效 DNA 中和第二种 mRNA 杂交的百分比则代表了在第一种组织中不表达而在第二种组织中表达的基因。

通过这些方法和其它方法，表明一个组织中的丰富 mRNA 在另一个组织中往往仅作为稀少 mRNA 而存在或甚至不存在。一般来说。哺乳动物某一细胞类型中只有 10% 的 mRNA 序列是该细胞类型所特有的，另外的 90% mRNA 序列在某些其它的细胞类型甚至在所有的细胞中同样存在。这就是说，哺乳动物(其它高等真核生物也可能如此)各类不同的细胞中均有相同的一组基因在表达。这组基因数目约在 10 000 左右。由于它们功能对于每个细胞都必需。因此有时叫做持家 (house keeping) 基因；同时，不同细胞类型又往往有种类不多的只在该特定的细胞类型中表达的基因，这些基因有时又叫奢侈 (luxury) 基因。尚不知在各种类型细胞中表达的奢侈基因的总和，有人估计可能超过持家基因的数目(10 000)，但不会超过它的 2～3倍。

由于持家基因和奢侈基因在各种细胞类型中的功能、表达范围和表达强度不同，因此，细胞对这两大类基因调控的途径和机制也可能不同。在个体发育过程中，表达的基因数目逐步减少。例如，海胆卵中含有 18 000 种 mRNA，在囊胚阶段，剩下 13 000 种，在原肠胚阶段为 8 500 种，到了幼虫阶段，只剩下 7 000 种 mRNA。更有甚者，在某些高度特化的组织中，表达的基因数目就更少了，例如成体海胆的管足和肠等组织中，表达的基因数目只有 2 500～3 000 种了。实验表明，胚胎后期表达的基因，其大多数在胚胎早期也是表达的。

### 五、胚胎分化过程中有可能存在着相对简单的分子开关

果蝇的某些突变能够引起严重的发育紊乱。例如，有一类显性突变，称为触角脚突变 (*Antennapedia*) 能够使果蝇头上触角部位长出脚来。这种脚与正常的脚形态相同，但生长的位置却完全不同。这种现象称为同形异位现象(homeosis)。引起同形异位现象的突变则称为同形异位突变(homeotic mutations)，如触角脚突变就是一类同形异位突变。在触角脚这类同形异位突变中，有些效应较弱，只引起触角的末端部分转变为脚的相应末端部分(爪和跗)，而其余部分仍然与触角相同。这种突变称为芒脚突变(*Aristapedia*)。许多触角脚突变都有可见的染色体异常现象(转位、倒位、缺失)，它们均出现在第 3 号染色体右臂的 84 B 带。

触角脚突变的存在，说明了少数基因的产物可以控制触角-眼睛这一胚盘中相关细胞的分化发育；芒脚突变则说明了不同的调节蛋白质或者几种调节蛋白质的不同组合可以控制不同的细胞亚群的分化发育。这些调节蛋白质就是生物体分化发育的分子开关。在果蝇中，有两

大同形异位基因簇。一个叫ANT-C(antennapedia complex)，包括*Dfd*,*Scr*,*ftz*和*Antp*四个基因；另一个叫 BX-C(bithorax complex)，包括*Ubx*,*abd-A*,*abd-B*三个互补群，每个互补群包含一个以上的基因。上述的同形异位基因都含有高度保守的180个核苷酸的DNA序列，编码羧基末端的60个氨基酸。这高度保守的180个核苷酸的DNA序列就叫做同形异位框(homeobox)。后来人们发现非洲爪蟾以及其他高等真核生物包括人类都存在着同形异位框。这些同形异位基因中有不少在胚胎发育过程中也起着重要的作用。现在已经知道，这些同形异位基因所编码的蛋白质都是区域特异性的转录因子。这些转录因子直接或间接地影响到控制胚胎发育的基因。果蝇的ANT-C和BX-C所影响的各个部位如图9-4所示。而ANT-C中的*ftz*基因则是体节分化基因的一个成员；如果*ftz*基因突变，则引起“偶数”体节(T2,A1,A3,A5,A7)的后半部和“奇数”体节(T1,T3,A2,A4,A6,A8)的前半部的缺失。这类转录因子中的同形异位框与DNA序列的相互作用机制将在第十章中讲述。

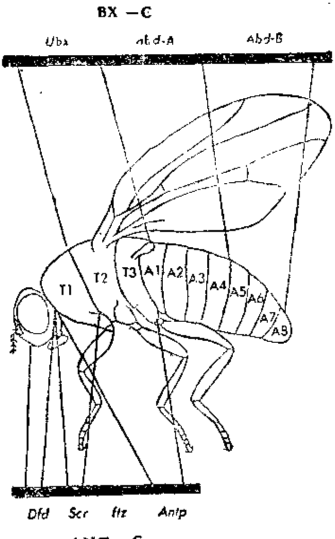



图9-4 果蝇的ANT-C和BX-C所影响到的各个部位。ANT-C中的*ftz*基因则作为体节分化基因的一个成员而起作用

## 第二节 DNA水平的调控[7~19]

### 一、基因丢失

在细胞分化过程中，可以通过丢失掉某些基因而去除这些基因的活性。某些原生动物、线虫、昆虫和甲壳类动物在个体发育中，许多体细胞常常丢失掉整条或部分的染色体，只有将来分化产生生殖细胞的那些细胞一直保留着整套的染色体。

马蛔虫受精卵细胞内只有一对染色体($2n=2$)。这对染色体上排列着许多着丝点，在由受精卵发育为成体的早期阶段只有其中一个着丝点行使功能，保证了正常有丝分裂的进行。当个体发育进行到一定阶段后，在将来分化产生体细胞的细胞中这对染色体破碎，形成许多小的染色体，其中有的小染色体含有着丝点。它在以后的细胞分裂中都将保持下去；而有些小染色体不含着丝点，它们因不能在细胞中正常分配而丢失。在将来形成生殖细胞的细胞中，不存在染色体破碎现象。

目前在高等真核生物(包括动、植物)中尚未发现类似的基因丢失现象。由于许多生物各类不同的细胞或细胞核都具有经过去分化和再分化而发育出完整个体的潜在能力，说明这些

细胞核内保存了个体发育所必需的全部基因。这种能力称为全能性(totipotency)

### 二、基因扩增

基因扩增指细胞内某些特定基因的拷贝数专一性地大量增加的现象，它是细胞在短期内为满足某种需要而产生足够的基因产物的一种调控手段。这方面的例子研究得最清楚的有两栖类和昆虫卵母细胞 rRNA 基因(rDNA)的扩增。

非洲爪蟾体细胞中 rDNA 拷贝数约有 500 个，而在卵母细胞中的拷贝数约为 2 000 000 个，增加了 4 000 倍。这些 rDNA 约占了整个卵母细胞 DNA 的 75%，可以用来转录合成卵裂期所需要的 $10^{12}$ 个核糖体。未扩增的 rDNA 成簇存在，重复串联在一起，形成核仁组织区。扩增后，新形成的 rDNA 并不是以一条长链 rDNA 串联重复的形式出现，而是以一个个环状 DNA 小分子的形式出现。这些小分子含有的 rDNA 拷贝数多少不等。绝大多数含有两个以上的拷贝。试验表明，和基因组 rDNA 中各个拷贝之间的间隔区的大小不一正相反，同一种环状 DNA 分子中各拷贝的间隔大小完全一致；但不同环状 rDNA 分子的间隔区大小各不相同。这些情况表明，同种环状 rDNA 分子可能都来自基因组上的一个 rDNA 拷贝。rDNA 扩增的分子机制还不太清楚，一般认为在扩增中基因组上的一个 rDNA 重复单位（包括转录区和间隔区）被切除下来且两头连接成环，然后再以这个环状 rDNA 单位为模板，进行滚环复制(见第三章)。应该指出的是，基因组中不同 rDNA 重复单位的扩增频率是不同的，有的重复单位不扩增。扩增一般在卵细胞成熟时停止，所合成的 rDNA 开始逐渐降解。当受精卵经过几轮分裂产生数百个细胞后，这类染色体外的 rDNA 便完全消失。

除上述在正常情况下因发育需要而出现的基因扩增外，外界环境因素也可能造成基因扩增。例如，二氢叶酸还原酶(DHFR)的抑制剂氨甲喋呤(methotrexate) 处理离体培养的细胞系，可以使这个酶结构基因 *dhfr* 扩增达到 40～400 个拷贝。于是，细胞产生更高的酶活性来增加对氨甲喋呤的抗性。进一步分析发现，通过 *dhfr* 基因扩增而抗氨甲喋呤的细胞系可分为两大类：一类是稳定系 (stable lines)，无论是否去除氨甲喋呤，扩增的 *dhfr* 基因都依旧存在；另一类是不稳定系(unstable lines)，扩增的 *dhfr* 基因在去除氨甲喋呤后逐渐消失。

稳定系中扩增的 *dhfr* 基因成簇串联在一起，位于一条染色体原来的 *dhfr* 位置上，而其同源染色体的 *dhfr* 则通常不扩增。扩增的*dhfr* 位置在染色时表现为均一染色区(homologeously staining region，简写作 HSR)，这可以在显微镜下观察到。与之相反，非稳定系中扩增的 *dhfr* 基因在基因组染色体外以所谓双小染色体 (double-minute chromosomes) 的形式存在。每个双小染色体含有2～4 个 *dhfr* 基因。双小染色体可以自主复制，但无着丝点，故在有丝分裂中不能均等地分配到子细胞中去，并常常因此而丢失。只有在选择压力(氨甲喋呤）存在时，双小染色体才能得以保持，否则不含双小染色体的细胞将很快繁殖而占主导地位。

当细胞接触氨甲喋呤时，最初产生非稳定 *dhfr* 基因扩增细胞株系，然后以一种尚不明确的机制产生在染色体上出现扩增基因的稳定株系。那么最初的非稳定基因扩增又是如何产生的呢？图 9-5 所示为对这一问题的解释。在 *dhfr* 基因附近出现局部 DNA 复制，复制后，含有 *dhfr* 的 DNA 片段以一种类似于重组的方式圈出而脱离染色体。几个含有单个 *dhfr* 的环状 DNA 可以通过重组而连成一个大的含有几个基因的拷贝的双小染色体。很有可能在正常情况下也不断出现 *dhfr* 基因扩增，只是因为没有氨甲喋呤的选择压力而趋于丢失。不过目前

发现有类似扩增现象的基因并不多。可能这些基因位于染色体的特殊部位，使它们有可能局部复制并圈出脱离染色体。

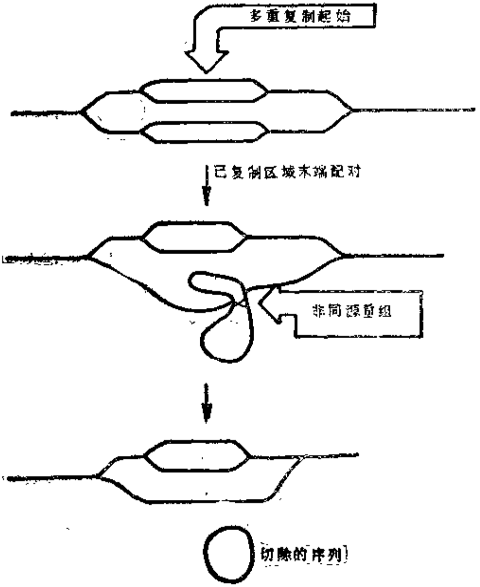



图 9-5 含 *dhfr* 基因的双小染色体的产生机制

### 三、基因重排

DNA 水平基因表达调控的另一个途径是基因重排。啤酒酵母 *Saccharomyces cerevisiae* 接合型的决定即属此类。单倍体酵母有a 和 $\alpha$ 两种接合型，分别由 MATa 和 MAT$\alpha$ 控制，这两个位点互为等位。一个特定的单倍体组细胞或是 a 接合型，或是 $\alpha$ 接合型。两个不同接合型的细胞可以接合，而相同接合型的细胞不能接合。对不同接合型细胞的识别是由于分泌激素的不同：MAT$\alpha$ 细胞产生 $\alpha$ 因子；MATa 细胞产生 a 因子。有趣的是，$\alpha$ 型可以转变成 a 型，a 型也可以转变成$\alpha$型。这种现象叫做接合型互变(mating-type interconversion)。接合型互变需要另一个位点上显性等位基因 HO。从 $\alpha$ 型和 a 型细胞可以互变而知，一个单倍体细胞中同时存在着 MAT$\alpha$ 和 MATa 的遗传信息，但在一个特定时刻只有其中的一个表达。倘若正在表达的 MAT 位点发生突变，该突变位点可以被野生型的MAT$\alpha$或 MATa 所取代。这说明正在表达的 MAT 位点以外的地方还分别存在着MAT$\alpha$ 和MATa 位点。现在知道，这两个位点和表达的 MAT 位点位于同一染色体上。一个位于 MAT 的左面，叫做HML$\alpha$，它是使细胞由 a 型转换为 $\alpha$ 型所必需的。另一个位于 MAT 的右边，叫做 HMRa，是由 $\alpha$ 型转变为 a 所必需的(图 9-6)。

解释接合型互变的模型叫做匣子模型 (cassette model)。如图 9-6 所示，在 MAT 位置

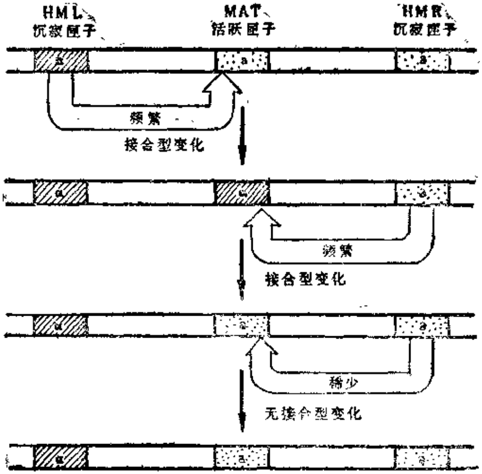



图 9-6　酵母接合型互变的匣子模型

上存在着活跃匣子。当这个匣子是 MATα 时，细胞就是 α 型的；当这个匣子是 MATa 时，细胞就是 a 型的。而在两个 HM 位置上存在着沉寂的匣子，左边的为 HMLα，右边的为 HMRa。处于 HM 位置上的与活跃匣子类型不同的沉寂匣子可以取代活跃匣子从而改变细胞接合型，而且在取代的同时，在原来的 HM 位置仍然保留一个拷贝的沉寂匣子。这就是说，在取代的过程中必然伴随着 DNA 的复制(详见第十一章第四节)。

序列分析表明，HMLα，MATα，MATa 均由 W，X，Y，Z1，Z2 五部分组成，而 HMRa仅由 X，Y，Z1 三部分组成(表 9-1)。可能在酵母基因组的进化过程中，MATα 首先占据决定接合型的位置。两种活跃匣子的差异只在 Y 区域，因此两种活跃匣子编码的基因产物也不同。然而，在 Y 区域的左端，四种匣子都有序列完全相同的启动子，两个开放阅读框 (ORF) 相背而行。那么，为什么活跃匣子(MATα 或 MATa)能够进行活跃的转录而 HMLα 和 HMRa 却保持沉默呢？

表 9-1　酵母四种与接合型相关的匣子的序列组成及长度 (bp)

| 匣　子 | W | X | Y | Z1 | Z2 | 总　长 |
|---|---|---|---|---|---|---|
| HMLα | 723 | 704 | 747 | 239 | 88 | 2 501 |
| MATα | 723 | 704 | 747 | 239 | 88 | 2 501 |
| MATa | 723 | 704 | 642 | 239 | 88 | 2 369 |
| HMRa | | 704 | 642 | 239 | | 1 585 |

人们在酵母基因组上发现了互不连锁的负责阻遏沉寂匣子基因表达的四个阻遏蛋白或其的基因 *sir1*，*sir2*，*sir3*，*sir4*。这种阻遏蛋白通过反式作用控制沉寂匣子的表达。然而，通过缺失突变的研究，这个阻遏蛋白的结合位点位于启动子上游 1 500 bp 以外的 E区。两个沉寂匣

子的E区序列完全相同，而活跃匣子却没有这样的结合位点。Sir 结合于离启动子如此之远的位点，它是如何能控制转录的问题至今还没有完美的答案。除了通过噜噗机制（第十章）影响 RNA 聚合酶的结合之外，还可能使沉寂匣子启动子处的染色质浓缩而改变局部染色体上基因的可转录性。虽然所有的匣子都含有对 DNA 酶 I 的超敏感点，但分布位置不同。在 MAT 匣子上，两个转录单位的上游区域各含有一个超敏感点，而这些敏感点在 HML 和 HMR 上不存在。DNA 酶 I 超敏感点被认为与基因表达有密切关系（见本章第三节），它们的不同分布可能是造成活跃匣子和沉寂匣子表达活性不同的原因之一。

另一种熟知的以基因重排来调节不同基因表达的例子是哺乳动物免疫球蛋白各编码区的连接。免疫球蛋白包括两条相同的轻链和重链。轻链包括可变区，恒定区以及二者之间的连接区。每个区域都由位于同一染色体不同位置的 DNA 片段编码，在形成活性基因前，一个可变区基因和一个恒定区基因及其上游的某一个连接区通过染色体内重组而连到一起。对于重链，除上述可变区，恒定区和连接区外，还包括一个介于可变区和连接区之间的歧化区，歧化区由同一染色体上另一 DNA 片段中几个串联的 D 编码区之一编码，重链活性基因的形成涉及到将这些编码区通过染色体内重组而连接起来。除可变区外，歧化区和连接区也是增加特异性免疫球蛋白数目的因素。不同的可变区，歧化区和连接区的连接使细胞具有产生多达几百万种免疫球蛋白的潜在可能性。本书限于篇幅，对于免疫球蛋白的基因及其表达不再作进一步的叙述。

## 第三节　活跃转录基因的染色质结构[20~35]

### 一、转录基因的核小体结构

真核生物染色体的基本结构单位是核小体。166～168 bp 的 DNA 以左手方向在组蛋白八聚体上缠绕两圈，一个分子组蛋白 H1 结合于核小体的 DNA 进出口处从而稳定了核小体结构。而核小体之间的连接 DNA 的长短随物种不同而不同，从几个 bp 直到大约 70bp 不等。用微球菌核酸酶进行有限的水解，可以得到长短不一的寡聚核小体的混合物。每个寡聚核小体含有的 DNA 是单个核小体 DNA 加上连接 DNA 长度的整数倍。进一步用微球菌核酸酶水解，就可以得到不含组蛋白 H1 和 146～148 bp（1.75 圈）的DNA。核小体纤丝的进一步缠绕就形成 6 个核小体一圈直径 30 nm 的螺线管状结构。这种结构就是染色质丝。30 nm 的染色质丝的稳定性依赖于组蛋白 H1 的存在。

活性基因与非活性基因是否也反映在核小体结构上呢？回答是肯定的。凡是活跃转录的基因全部位于常染色质中，处于异染色质中的基因则不表达。然而并非所有位于常染色质部位的基因都在进行活跃的转录。在一个特定的时间内，只有一部分常染色质的基因在表达。在转录时，RNA 聚合酶和大约 50 bp 的 DNA 结合，并且其中有 12 bp 要发生解链。再说，RNA 聚合酶（500 kDa）要比一个核小体（262 kDa）大得多。很难设想转录过程是 RNA 聚合酶沿着核小体表面的 DNA 链进行的。相反，转录过程必然涉及到核小体结构的变化。

rRNA 基因在大多数情况下都是进行活跃的转录。通过电子显微镜的观察，rRNA 基因的实际长度与电子显微镜下观察到的长度之比仅为 1.2，这说明在 rDNA 上基本没有核小结构。位于两个 rDNA 转录单元之间的间隔区也处于高度松弛状态，包装比也仅为 1.4。因此，在整个核仁区内，绝大部分 DNA 并不组织在核小体中，而是以游离状态存在。

另一方面，在SV40的微染色体中，正在转录的基因仍然有核小体结构。这就说明，转录并不要求整个基因都处于无核小体状态。这可能与转录频率有关。用微球菌核酸酶对染色质进行处理，然后进行凝胶电泳。结果发现转录的基因和不转录的基因都出现间隔为200 bp的DNA梯带。这说明转录的基因和非转录基因的核小体数目没有显著的差异，只是转录基因对微球菌核酸酶更为敏感，降解得更快，也就是说其核小体数目略少一些，很可能是当RNA聚合酶经过的部分的核小体结构中的组蛋白被置换。当RNA聚合酶离开后立即又恢复了核小体结构。近来人们用SP6 RNA聚合酶和RNA聚合酶Ⅱ对含有核小体结构的DNA进行的离体转录试验表明，这两种RNA聚合酶都不能从有核小体结构的启动子上发起转录，但可以进行链的延伸并同时置换核小体。

## 二、DNA酶Ⅰ优先敏感性和HMG蛋白质

用DNA酶Ⅰ处理染色质以得到酸溶性的DNA片断。处理时间越长，产生的酸溶性的DNA比例就越大，直至全部染色质都被降解掉。用特定的cDNA或mRNA作探针。可以测定某个基因被DNA酶Ⅰ降解情况。试验表明，当总体DNA的10%被降解成酸溶性物质时，50%以上活跃表达基因已优先降解了。若把从鸡红细胞中提取出的β-珠蛋白基因和卵清蛋白基因分别作上述处理，β-珠蛋白基因很快就被降解，而卵清蛋白基因的降解程度则很小。相反，若从鸡输卵管细胞中提取这两种基因并作同样处理，优先降解的是卵清蛋白基因而不是β-珠蛋白基因。这些结果说明活跃基因比不活跃的基因更优先受到DNA酶Ⅰ的攻击。当采用代表整个细胞各种类型mRNA群体的cDNA作探针进行上述试验时，不管是转录速率极高的丰富mRNA基因还是转录速率较低的稀少mRNA基因都表现出对DNA酶Ⅰ的优先敏感性(preferential sensitivity)(虽然在敏感程度上存在着差别)。由于稀少mRNA的基因在一个特定时刻所结合的RNA聚合酶数目很少，上述结果说明对DNA酶Ⅰ的优先敏感性不是源于转录过程中RNA聚合酶的作用，而是可转录基因的一个基本特征。

对DNA酶Ⅰ的优先敏感性不仅仅表现在活跃基因的转录区，而且也表现在活跃基因两侧的DNA区段上，可以把活跃基因转录区及其两侧的DNA酶Ⅰ优先敏感的DNA序列定义为一个区域(domain)。鸡卵清蛋白基因的DNA酶Ⅰ敏感区域约有100 kb长，这一区域内转录区和两侧的非转录区对DNA酶Ⅰ同样敏感。鸡β-珠蛋白基因簇和它的5′端约6～7 kb和3′端约8 kb组成一个DNA酶Ⅰ敏感区域，这一区域的非转录区对DNA酶Ⅰ的敏感性介于转录区和其它DNA区域之间。

DNA酶Ⅰ优先敏感区域的发现表明，活跃基因上游和下游序列也存在着染色质结构的变化。由于这一变化的范围很广，包括5′端启动子本身及其上游部位。因此很难用RNA聚合酶转录时的作用来解释。所以可以认为染色质结构变化的确发生在具体的转录过程之前。

究竟是什么因素造成了DNA酶Ⅰ的优先敏感性呢？用DNA酶Ⅰ处理染色质时，优先释放出两种非组蛋白，它们是HMG14和HMG17。HMG为high-mobility group的缩写，即高度可动性蛋白质，为一组分子量不大的非组蛋白。它们可被低盐溶液(0.35mol/L NaCl)抽提，说明它们和染色质其它组分的结合较松弛。把HMG14和HMG17从鸡红细胞染色质中提取出来后珠蛋白基因不再表现出对DNA酶Ⅰ的优先敏感性，而在盐抽提过的染色质中加入这些蛋白质后，敏感性又可恢复。不过若把这些蛋白质加于盐抽提过的鸡脑细胞染色质中后，其中的珠蛋白基因并不表现出对DNA酶Ⅰ的优先敏感性。这说明这些HMG蛋白质不是造

成优先敏感性的唯一原因。红细胞和脑细胞中一定还存在着其它因子的差异，这些因子在盐抽提后仍留存在染色质中。这些因子决定了红细胞的珠蛋白基因在HMG 14 和 HMG 17 存在时对DNA 酶 I 敏感。

HMG 14 或 HMG17 单独存在时也有同样效应，这说明 HMG14 和 HMG17 作用类似。大约一分子的 HMG 能使含有 10～20 个核小体的染色质对 DNA 酶 I 优先敏感。进一步的分析发现，HMG 14 和 HMG 17 的效应不但使完整的染色质区段对 DNA 酶 I 敏感，而且对经微球菌核酸酶降解产生的核小体单体甚至核小体核心颗粒也有同样的效应。这说明活跃基因的核小体核心颗粒一定具备某些使它们和 HMG 14 或 HMG 17 结合或相互作用的特点。这些特点在非活跃基因中不存在。

DNA 酶 I 优先敏感状态的获得和消失与基因的活跃转录在时间上有什么关系呢？5 日龄鸡胚胎的红细胞样细胞 (erythroid) 只产生胚胎型珠蛋白。但它的胚胎型珠蛋白基因和成体珠蛋白基因都对 DNA酶 I 优先敏感，同样，只产生成体珠蛋白的 14 日龄鸡胚胎细胞中的这两个基因也都表现出对DNA酶 I 敏感性。不过对这一结果可以有两种解释：一种认为 DNA 酶 I 敏感状态出现在基因实际作为模板而转录之前，可以认为这种敏感状态是“可转录基因”的一种标记；另一解释是，既然 DNA 酶 I 敏感区域包括基因两端的上、下游序列，由于胚胎型珠蛋白基因和成体珠蛋白基因的连锁成簇，其中一个基因的结构活化会波及另一个基因的结构，使两者同时获得 DNA 酶 I 优先敏感性。目前有充足的证据证明，DNA 酶 I 优先敏感性在基因活性消失后仍然存在。如成熟红血球中无新的珠蛋白 mRNA 转录产生，但珠蛋白基因仍保持着 DNA 酶 I 优先敏感状态。目前尚不清楚一个基因在何时以何种机制从 DNA 酶 I 敏感状态转变为不敏感状态。

综上所述，DNA 酶 I 优先敏感状态的出现是转录所必需的，但它并不是只在转录进行时才存在，所以从这个意义上讲它又是与转录过程本身相互独立的。可以认为，DNA 酶 I 优先敏感状态是转录的必要条件而不是充分条件。

## 三、DNA 酶 I 超敏感点

如用极低浓度的 DNA 酶 I 处理染色质，切割将首先发生在少数特异性位点上。以果蝇为例，整个基因组可以被切成一系列大小为 2 kb～20 kb 的特异性片断，这些特异性切点叫做DNA酶 I 的超敏感位点 (hypersenstitive sites)。超敏感位点的存在是染色质的特点，若用游离DNA 作底物则无超敏感位点表现。

超敏感位点的位置可以用双重切割法分析。将染色质用极低浓度 DNA 酶 I 切割，提取DNA，再用某种限制性内切酶切割，然后走凝胶电泳，把双重切割的产物按分子量大小分开，再用代表着所用限制酶切点附近序列的探针(mRNA 或 cDNA)进行分子杂交。凝胶上出现的杂交带即代表了从限制酶切点到 DNA酶 I 超敏感切点之间的序列，可以从这段序列的长度来得出这两个切点之间的距离从而定位 DNA 酶 I 超敏感切点。通过对大量基因进行的试验，发现每个活跃基因都有一个或几个超敏感点，大部分位于基因 5′ 端启动子区域，少部分位于其它部位如转录单元的下游。当一个基因附近有着几个超敏感位点时，其敏感程度有一定差异。5′ 端的超敏感点只出现在基因正在活跃表达的细胞中，在该基因不表达的细胞中不存在。例如，对于鸡$\beta$-珠蛋白基因簇，胚胎阶段超敏感点出现在胚胎型基因而不是成体型基因的5′ 端；而在成体阶段情况则正好相反，超敏感点出现在成体型而不是胚胎型 $\beta$-珠蛋白基因的

5′端。这说明超敏感点和基因活性有一定关系。果蝇唾液腺中编码胶蛋白的 *sgs4* 基因转录区上游 330 bp 和 405 bp 左右分别有一个超敏感点。有些突变体缺失了包括这两个超敏感点的大约 100 bp,虽然转录起始点上游区域的将近 300 bp 仍然存在，这些突变体却失去了合成胶蛋白的能力。由此可见,这些超敏感点是 *sgs4* 基因活跃状态所必需的。

超敏感点的出现和存在与转录在时间上有什么关系？目前已有的实验证据表明，超敏感点建立于转录开始之前。然而,超敏感点一旦建立起来,除去诱导物后转录便很快停止,而超敏感点并不立即消失。由此可见,超敏感点的建立与维持有着不同的机制,同时超敏感点的存在与转录作用并非严格同步。这也说明,超敏感点的存在只是转录起始的必要条件,而不是充分条件。

超敏感点的性质是什么呢？现在知道每个超敏感点实际上并不是某个特定的碱基位置，而是一段长约 100～200 bp 的 DNA 序列。在这个序列中,包括许多核酸酶(包括微球菌核酸酶、DNA 酶 I 、 II 和某些限制性内切酶)的优先敏感点。鸡成体型 $\beta$-珠蛋白基因 5′端的超敏感点位于约 −50 至 −280 bp 之间,这一区域包括可被微球菌核酸酶，DNA 酶 I 和限制酶 Msp I 优先切割的位点。进一步分析表明，至少其中的一部分序列以游离 DNA 的形式存在。疱疹病毒胸苷激酶基因的 5′端超敏感点位于 −4～−182 bp 之间，这一区域也可被上述的核酸酶优先切割。试验还发现,活跃转录的鸡珠蛋白基因的超敏感点可被只切割单链 DNA 的S1 核酸酶切割，表明超敏感点部位的 DNA 状态与其它部位不同，很可能有局部的解链存在。综上所述，超敏感点可能反映了染色质的某种结构改变，在超敏感点区域内，核小体结构或是根本不存在或有不同寻常。其第一个线索是 SV 40 的微染色体上，在复制原点和晚期 RNA 转录起点之间，缺少核小体结构(图 9-7)。

图 9-7 SV40 微染色体的电子显微镜照片，在其复制原点附近缺乏核小体结构

超敏感点与它下游转录起始点之间的距离依不同的基因而异。在果蝇 *sgs4* 基因中这一距离至少为 250 bp，鸡 $\beta$-珠蛋白基因的超敏感点恰好在转录起始点的上游；疱疹病毒胸苷激酶基因的超敏感点大概占据了整个启动区域。超敏感点的出现对基因激活作用的机制仍是个谜，可能它改变了基因的启动子及其两端区域的染色质或 DNA 状态，使 RNA 聚合酶得以结合并起始转录。

有迹象表明，活跃基因上DNAase超敏感位点的出现可能与不同位点上调节蛋白质的相互作用而形成的DNA噜噗结构有关(见第十章)。

## 四、组蛋白的修饰作用

真核生物的组蛋白上某些特定的氨基酸残基往往被共价修饰。修饰作用包括乙酰基化、甲基化和磷酸化。乙酰基化和甲基化发生在赖氨酸残基的自由氨基上，有时甲基化还发生于精氨酸和组氨酸残基。磷酸化发生在丝氨酸的羟基及组氨酸上。所有这些修饰作用都有一个共同的特点，即降低组蛋白所携带的正电荷(即碱性程度)，因此，可以通过凝胶电泳而把未经修饰过的组蛋白区别开来。不过，上述的这些修饰作用都只发生在组蛋白的内部氨基酸上，而且都是暂时性的。与此相反，大多数组蛋白的N端在合成时就有稳定的乙酰基。

组蛋白修饰作用一般在细胞周期的某一时刻发生，而在另一时刻消失。由于它使正电荷减少，故被认为可能与组蛋白功能的改变有关。不过到目前为止尚未发现组蛋白修饰作用与染色质的功能之间的明确关系。

鳟鱼精巢组蛋白(实际上这里的组蛋白被鱼精蛋白所取代)的H3和H4，是乙酰基化和去乙酰基化的主要靶点(各有4个修饰点)，H2A和H2B较少受影响。不同鳟鱼个体组蛋白的修饰点很少相同。在同步培养的细胞中，原来存在的以及新合成的核心组蛋白于S期(即组蛋白和DNA的合成期)被乙酰基化和甲基化，接着修饰基团又被去掉。这些修饰作用与染色质复制同时发生，表明修饰作用可能和核小体的装配有关——组蛋白正电荷的减少会暂时松弛它们与DNA的结合，便于更好地控制核小体装配。

乙酰化存在的短暂性给研究带来了一定的困难，这可以通过向生长中的细胞加入丁酸来克服。丁酸抑制组蛋白脱乙酰酶，使所有的核心组蛋白乙酰基化。伴随着乙酰基化，染色质出现类似于活跃基因的特点，对DNA酶I和微球菌核酸酶更加敏感，不过尚未发现生物体内活跃基因和不活跃基因之间乙酰基化程度的差异。所以，虽然乙酰基化对染色质结构的确有影响，它的生物学意义仍是个谜。

磷酸化和去磷酸化发生在组蛋白H1上，但具体时间和核心组蛋白的修饰不同。哺乳类动物离体细胞H1的磷酸化主要发生在有丝分裂期，每个H1分子可接受6个磷酸基。这些磷酸基在子细胞核形成时减少了80%。由于H1是染色质由10nm纤维产生30 nm纤维所必需的，它的磷酸化在有丝分裂时的出现被认为可能与染色质的收缩有关。某些不能进行H1磷酸化的温度敏感突变体中，染色质复制受阻，因而不能进行细胞分裂。这一观察支持了上面的推论，但不知H1磷酸化是否是染色质收缩的主要原因。

真核生物组蛋白还有一种特殊的修饰作用，即组蛋白H2A上连接一个叫做泛素(ubiquitin)的蛋白质，这种修饰作用最初在鼠肝中发现。当时把修饰过的H2A叫做A24，现在把它叫做uH2A。泛素广泛存在于从酵母到哺乳类动物的各类生物中，它含有76氨基酸残基，以其C末端甘氨酸的羧基和H2A上第119残基(H2A共有129个残基)即赖氨酸上的自由ε-氨基形成异肽键(isopeptide bond)。泛素为酸性蛋白质，含有许多谷氨酸和天冬氨酸残基。它与H2A结合后形成uH2A使二者的酸碱性得到中和。

真核细胞中大约5～15%的H2A以uH2A的形式存在，一般每个组蛋白八聚体中只有一个H2A带有泛素，故大约10～30%的核小体是泛素化的。此外，很小一部分H2B也可被泛素化。

果蝇中大约20%的核小体含有uH2A，可以把这些核小体的DNA提取出来用不同的探针作分子杂交进行分析。试验表明，大约50%的活跃基因的序列处于泛素化的核小体中，而这些核小体中只含有不到4%的卫星DNA序列。因此uH2A基本上不存在于异染色质内，而主要位于活跃基因序列的核小体中。

关于泛素化的组蛋白是否与转录激活有关，现在仍未定论。有人发现转录的抑制不影响uH2A的总含量。然而，uH2B可能与转录激活有密切关系。

## 五、DNA的甲基化和去甲基化

真核生物DNA中大约2～7%的胞嘧啶(C)存在着甲基化修饰，其中卫星DNA序列最为强烈，其它序列也广泛含有甲基化C。绝大多数甲基化发生在CG二核苷酸对(CG doublet)，大多数CG二核苷酸对都是甲基化的，而且两个C通常都发生甲基化，即：

5′……$^m$CpG ……3′

3′…… GpC$^m$……5′

但有时只有一个C是甲基化的，这种CG二核苷酸对叫半甲基化的(hemimethylated)。采用特异性切割CG二核苷酸对的限制性内切酶大大便利了对甲基化的研究。有两种这样的限制酶：HpaⅡ识别并切割未甲基化的CCGG序列，但对甲基化的CG二核苷酸对则不起作用；而MspⅠ可以识别并切割所有的(包括甲基化的和未甲基化的)CCGG序列。因此，可以用MspⅠ来鉴别CCGG的存在，用HpaⅡ来鉴别这些CCGG序列是否甲基化(图9-8)。

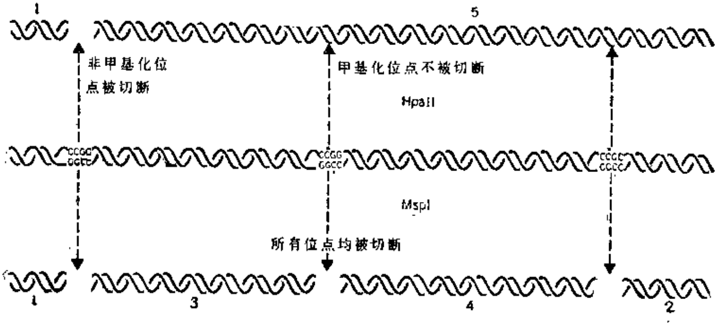



图9-8 限制性内切酶HpaⅡ和MspⅠ对甲基化及非甲基化CCGG序列的切割
*＝甲基化碱基

应用上述限制酶对许多生物的不同DNA序列(包括蛋白质编码序列，rDNA和病毒DNA)进行研究。结果表明，在这些DNA序列的CCGG位置上，有的位置在同一生物所有的组织中都甲基化了；有的在所有的组织中都没有甲基化；但是，一小部分位置在该基因不表达的组织中是甲基化的而在该基因活跃表达的组织中是未甲基化的。分别把甲基化和未甲基化的病毒DNA或细胞核基因引入细胞，甲基化的不表现活性，而未甲基化的表现出基因活性。因此，可以把活跃基因的状态称为甲基化不足(under-methylation)。

一个活跃基因表现甲基化不足状态的范围究竟有多大呢？对鸡红细胞样细胞α-珠蛋白基因簇的研究可以回答这个问题。这里，甲基化不足位于α-珠蛋白基因簇(含两个成体α-珠

蛋白基因)中第一个基因上游 500 bp 到第二个基因下游 500bp 之间的 CG 二核苷酸对上。在这一范围内，甲基化不足不但出现在转录区，而且也出现在两基因间的非转录区中。有趣的是，甲基化不足的范围和 DNA 酶 I 优先敏感区域十分吻合。

在活跃基因中，只有一小部分甲基化 CG 二核苷酸对的去甲基化与基因激活有关，也许这些位置处于对基因表达关系十分密切的序列(如启动子)中，它们去甲基化对基因可转录状态的形成至关重要。例如在非洲爪蟾 rRNA 基因中，在胎胚发育时期，该基因上大部分甲基化 CG保持不变，只有靠近启动子的两个位点产生去甲基化。

同一组织不同细胞中甲基化在 DNA 上位置总是一致的。图 9-9 为对细胞分裂时甲基化

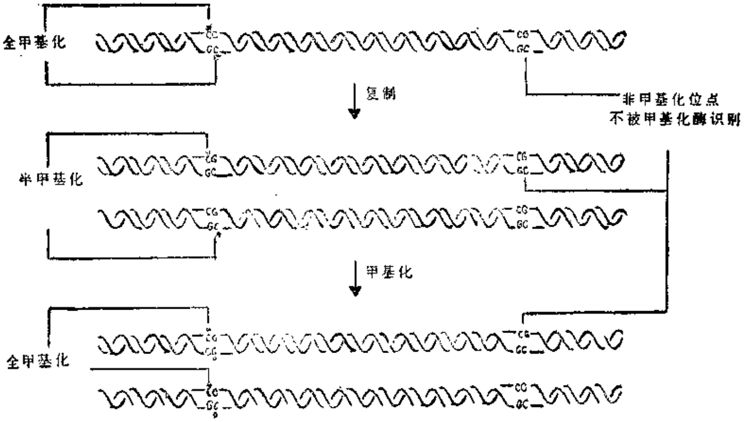



图 9-9 甲基化 CG 二核苷酸对的复制

*=甲基化碱基

位置维持不变的比较流行的一种解释。如果 DNA 上某一个 CG 二核苷酸对呈完全甲基化(即两条链都甲基化)，在 DNA 复制所形成的两个子体 DNA 分子中就会各有一个半甲基化 CG 二核苷酸对。DNA 甲基化酶可以识别半甲基化 CG 二核苷酸对并使它完全甲基化，试验中发现，原来甲基化的 CG 二核苷酸对可以在许多轮 DNA 复制后仍保持甲基化，而原来没有甲基化的 CG 二核苷酸对则不能被从无到有地甲基化，这一结果和上述解释是一致的。另外对于在体外使所有的 C 都甲基化的 $\beta$-珠蛋白基因 DNA，导入细胞核并复制后只有 CG 二核苷酸对的甲基化能得以保存而其它序列中C的甲基则失去，这进一步说明细胞核内甲基化酶的作用靶点是半甲基化的 CG 二核苷酸对。

去甲基化的途径可能有两种，或是由去甲基酶去除，或是在 DNA 复制时由于某种原因甲基化酶未能在半甲基化处催化甲基化，从而在以后的 DNA 复制中产生完全去甲基化的 DNA 分子。

采用 5-氮胞苷 (5-azacytidine)可以人为地造成去甲基化。5-氮胞苷可以在 DNA 中取代胞苷，它没有游离的 5-H，因而不能接受甲基。用 5-氮胞苷处理细胞，可以改变基因表达和细胞分化的状态。比如，使非肌肉细胞的前体分化成肌肉细胞，诱导整合在染色体中的原病毒表达，激活哺乳类雌性动物失活的X染色体上某些基因的表达。后一发现似乎表明 X染色体

的失活与 DNA 甲基化状态有关。

值得一提的是，DNA 甲基化/去甲基化与基因活性的关系并不是绝对的和普遍的，脊椎动物 DNA 的甲基化程度比较高，尤其在哺乳类更是这样；而在无脊椎动物 DNA 中，甲基化程度就低得多；有些生物(包括果蝇和其它双翅目昆虫)中至今未发现DNA甲基化。这说明甲基化不是一种原始现象，它可能出现于某一进化时期，并随进化程度的提高而逐步增强。因此，甲基化/去甲基化在基因活性调控中的意义依生物不同而有所不同。即使对于同一种生物，甲基化对不同基因的活性也有不同效应。非洲爪蟾精细胞 rDNA高度甲基化。但在注射进卵母细胞核中以后其转录速度不亚于完全未甲基化的 rDNA。可是，同时在爪蟾卵母细胞内，其它基因的转录却因甲基化而大受抑制。rDNA 和其它基因的这种不同可能反映了同一生物体有不同的方法来克服甲基化对不同基因的抑制作用。对于 rDNA，生物体在进化中形成了一种容忍它的甲基化存在的能力，能够在不发生去甲基化的情况下转录它；对于其它基因，生物体通过去甲基化的方式来调节对它们的转录。这种情况可能与转录所用的RNA聚合酶种类有关。

过去对 DNA 甲基化的研究基本上都是采用限制酶 Hpa Ⅱ 和 Msp Ⅰ。必须指出，这些酶所识别的 CG 二核苷酸对(存在 CCGG 序列中)只占全部 CG 对子的一小部分。这在很大程度上限制了对 DNA 甲基化作为一个整体的存在、产生、维持和消失的机制以及在基因调控中的意义的研究。目前已经在技术上找到了直接鉴定所有 CG 序列的研究方法。

## 第四节　转录水平的调控[1,6,36~65]

### 一、Britten-Davidson 模型

在 DNA 水平的调控，只是真核生物基因表达的一个次要和辅助手段。更多的基因调控发生在转录及其以后的阶段。在本章第一节我们谈到真核生物细胞中一般有数千至一万多个基因表达，不同类型的细胞中由不同组合的基因表达，产生特定的表现型。那么每种细胞类型是如何保证"正确"的基因组合的表达呢？这可以通过两种基本方式做到。

一种途径是基因重复。有些基因在基因组中有几个相同的或相似的重复拷贝。不同类型的细胞表达不同的拷贝。这里，每个拷贝处在与其细胞类型相应的特异性控制之下。现在已经知道了不少的基因家族，如珠蛋白基因家族等。其不同拷贝的表达一定处于不同的调控系统的控制之下。

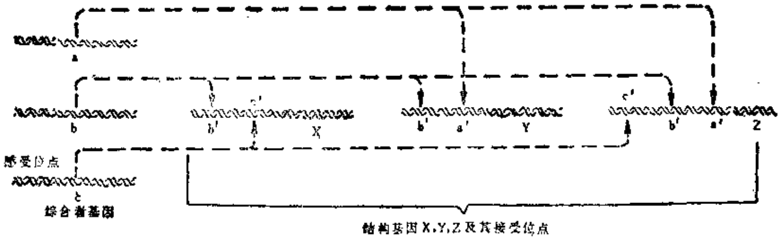
另一种途径是，对于单一拷贝的基因，可以通过复合控制系统来调控它的转录活性。在结构基因的5′端存在着可被控制因子识别的位点，与控制因子结合后，可以以某种方式促进转录。Britten 和 Davidson 提出了这种调控机制的模型(图 9-10)。在结构基因的 5′ 端连接有一段称为接受位点(receptor site) 的序列，它可被某种激活因子 (可能是 RNA 或蛋白质)所

图 9-10　真核生物基因表达调控的简单模型

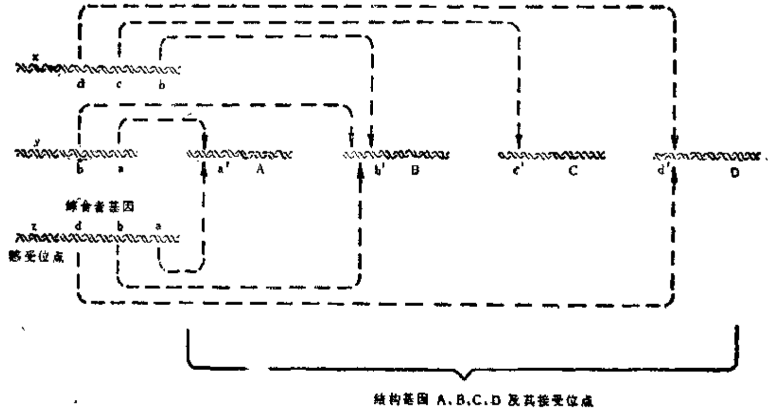
识别。激活因子由它的综合者基因(integrator)产生。这种情况类似于细菌操纵元与其调节基因之间的关系，但不同的是，真核生物的综合者基因自身还受到与它相邻的感受位点(sensor site)的控制。感受位点负责接受生物体对基因表达的调控信号(如激素等)。

根据上述模型，一个特定的激活蛋白可以同时控制含有相应接受位点的许多结构基因的表达。因此，从基因表达调控的意义上说，所有含有同样接受位点的基因组成一组(set)基因，类似于原核生物的一个操纵元或调控元。很可能同属一组的结构基因编码在功能上相互联系的蛋白质，比如同一个生化途径中不同的酶。

如果一个结构基因拥有几个不同的接受位点，每个接受位点可以被一个特异性的激活因子识别，这样它就可能作为不同组的成员而在不同的情况下表达。图 9-11(Ⅰ)描述了 X、Y、Z三个结构基因的组合表达情况，当综合者基因 a 产生激活因子 A 时，使 Y 和 Z 表达；当 b 产生出激活因子 B 时，X、Y 和 Z 均表达；当 c 基因产生出激活因子 C 时，X 和 Z 基因表达。




(Ⅰ)




(Ⅱ)

图 9-11 Britten-Davidson 关于真核生物基因表达调控的模型。

(Ⅰ) 一个结构基因受不同综合者基因—接受位点系统的控制。

(Ⅱ) 一个结构基因受不同感受位点的协调控制。

除了图 9-11(Ⅰ)所描述的一个基因拥有几个接受位点的情况外，基因表达的协调控制还可以通过感受位点控制不同的综合者基因而达到。如图9-11(Ⅱ)所示，在接受了特定控制信号的刺激后，一个感受位点可以同时产生几种激活因子进而同时激活几组不同的结构基因。同处于一个感受位点控制之下的所有结构基因叫做一套 (battery) 基因。像图 9-11(Ⅱ) 所示的那样，一组基因可以同时是不同套的成员。

根据上述模型，真核生物基因表达的协调控制是多级别的。一个特定的基因可以与其它任何的基因协调表达，上述模型要求有重复的综合者基因和接受位点，而重复序列正是真核生物DNA 中大量存在的。但必须指出的是，这个模型和生物体内实际存在的错综复杂的调控系统相比是过于理想化了，而且，它所提出的调控机制还必须经受实验的检验。就目前所知，已发现的激活因子基因的表达都是本底组成型表达，而非诱导合成，诱导只是使它具有促进转录的活性。

## 二、基因调控的顺式作用成分

从 70 年代后半期，人们利用定点突变的方法以及脚印法系统地研究基因转录的调控位点。高等真核生物的蛋白质基因表达的调控需要两种 DNA 序列，即启动子和增强子。启动子一般位于转录起始点上游 100bp 的范围内。启动子是 RNA 聚合酶进行精确而有效的转录所必需的；而增强子的作用在于增加转录的速率。增强子所具有的特性是：增强子可以在很远的距离作用于顺式连接的启动子，这种作用是没有方向性的。也就是说，增强子既可以位于启动子的上游，也可以位于启动子的下游。然而，近年来工作表明，在作用机制方面，增强子与启动子中调控转录的成分有不少相似之处。

通过对大量的启动子进行分子遗传学分析，发现蛋白质基因启动子的一般模式。大多数启动子在 −25 附近都有一个 TATA 框，在 −75 附近有一个 CCAAT 框（也有写作 CAAT 框的）。这两种调控成分我们已在第六章中作了初步的阐述。上游启动子成分 (upstream promoter elements，简称UPE)除了CCAAT框以外，还有其他各种因基因而异的上游启动子成分。如 GGGCGG，GCCACACCC，ATGCAAAT 等。现在已经知道，TATA 框是控制转录精确性的序列，而 UPE 则控制转录起始的频率。启动子的强度就决定于 UPE 的数目和种类。这些 UPE 序列相对于 TATA 框的方向不影响它们的作用。例如 GGGCGG 序列换成 CCGCCC 同样起作用。在 UPE 与 TATA 框之间插入整圈数 DNA 双螺旋和插入 n.5 圈双螺旋对转录的影响，前者要比后者小得多(见第十章)。

通过对$\beta$-珠蛋白基因的启动子进行密集的突变研究，已经发现该启动子上有一个TATA 框和两个 UPE序列，其中一个是CCAAT框，另一个是 GCCACACCC(图 9-12)。所有的 UPE 序列上都与特定的转录因子相结合，这可以通过脚印法和凝胶阻滞法进行检验。很多这些因子已通过识别位点亲和层析法加以提纯。

增强子同样是通过特定的转录因子与之结合而实现其对转录的增强作用。SV40 增强子是最早发现也是研究最多的一个增强子。它与大 T 抗原结合位点，TATA框，SP1 结合位点等调控成分构成极其复杂的调控区(图 9-13)，对 SV40 的早期基因的表达，晚期基因的表达以及 DNA 复制均有影响。这一调控区的作用机制也极为复杂，有调控蛋白质与 DNA 的相互作用，有结合于 DNA 上的调控蛋白质分子之间的相互作用，也有游离的调控蛋白质的相互作用从而影响它们与 DNA 的作用。由于调控蛋白质的结合，SV40 调控区缺乏核小体

结构。

GTAGAGCCACACCCTGGTAAGGGCCAA CTGCTCACACAG

−100 −80

GATAGAGAGGGCAGGAGCCAGGGCAGAGCATATAAGGTGA

(a) −60 −40

GGTAGGATCAGTTGCTCCTCACATTTGCTTCTGACATAGT

−20 +1 +10 +20

(b)

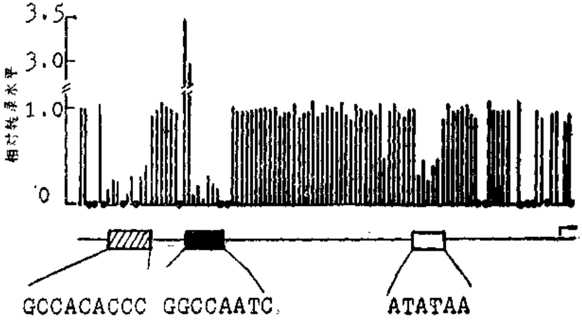



图 9-12 β 珠蛋白基因的启动子

(a) 启动子的 DNA 序列；(b) 通过突变效应的研究而发现的两个 UPE。

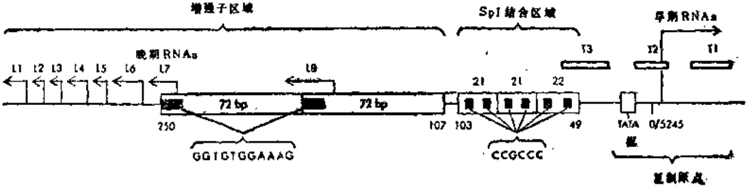



图 9-13 SV40 基因表达调控区结构示意图

在图 9-13 中，两个正向重复的 72 bp 序列（−107～−178，−179～−250）构成了增强子的主要部分。根据定点突变试验结果，SV40 增强子至少包含三个不连续的 15～20 bp 长的增强子成分（enhancer element），分别为 A，B，C（图 9-14）。这三个增强子成分能互相协同以促进转录。人工重组的两个或两个以上的相同或不同的增强子成分，也能增进转录。也就是说，增强子必须有两个或两个以上的增强子成分才具有自动的促进转录的功能。对于增强子成分作进一步的分析，发现有功能的增强子成分也都是由两个部分所组成，而且这两个部分必须紧密相连，明显地表现出对间距的敏感性，这种构成增强子成分的亚单元叫做增强子元（enhanson）。每个增强子元都是一个转录因子的结合位点。就目前所知，SV40 增强子主要有 coreA（−271～−263），coreC（−248～−240），TCⅡ（−239～−233），TCⅠ（−230～−224），SphⅡ（−216～−208），SphⅠ（−207～−199），P（−193～−185）等增强子元。coreA 的结合因子还没有纯化出来，人工构建的 coreA/coreA 增强子成分是有很高活性的，而 coreC/coreC 则没有活性。增强子成分 C 是由 coreC 和 TCⅡ所组成，它们分别与激活蛋白（转录因子）Ap3 和 Ap2 结合。TCI 与 Ap2 的结合能力较弱。此外 Ap2 还能结合于 SP1的

结合区以及早期基因的转录起始点附近。增强子成分B中的 SphⅡ和 SphⅠ有一个共同的转录因子。尽管其中 SphⅡ 和 SphⅠ 的交界处产生了八核苷酸序列（ATTTGCAT）的相似序列，但并不是由 NF-A 结合的，因为人工构建的增强子成分 SphⅡ/SphⅡ 的活性比天然的成分 B 更强，而 SphⅠ/SphⅠ几乎没有活性，但 SphI/SphⅡ却又具有天然增强子成分 B活性的 50%。这些都说明成分 B 是由两个强度不同的各自独立的增强子元组成的。它们与转录因子的亲和力不同，SphⅡ 对于转录因子的结合能力比 SphⅠ大得多。增强子元 P 是 AP1的结合位点，但实际表明，P 的缺失对于 SV40 增强子的功能没有什么影响。

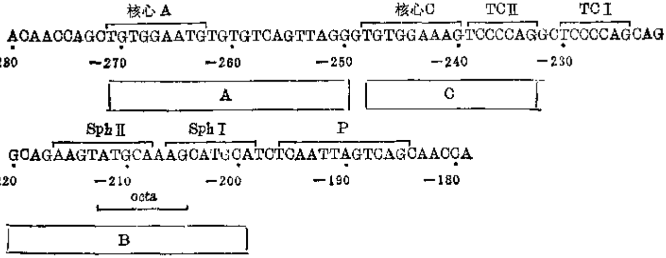



图 9-14　SV40 增强子中远上游重复序列（−179～−250）及其附近的增强子成分和增强子元

由图 9-14 可见，增强子成分 A，B，C，都是由两个增强子元构成的，而且这两个部分紧密相连，没有间隔。实验表明，人工组建的coreA/coreA 之间如果插入 1～2 个 bp，则活性只有无间隔的 coreA/coreA的6%；插入5个bp 则完全失去活性。实验还表明，尽管两个增强子元完全紧密相连，但它们与激活蛋白的结合仍然是各自独立的。用已纯化的Ap3和 Ap2 与增强子成分 C 的结合，完全是各自独立的，各占领成分 C 的一半。因而，增强子成分 C 是由两个不同的增强子元 coreC 和 TCⅡ 组成，它们分别与两个不同的转录因子 Ap3 和 Ap2 结合，从而生成一个能够促进转录的复合物。而增强子成分 B 是由两个基本相同的增强子元 SphⅡ和 SphⅠ组成，结合着完全相同的转录因子，也构成了能够促进转录的复合物。这些能够促进转录的复合物又称为能产型复合物(productive complex)。如果一个孤立的增强子元结合了一个结合蛋白或者一个增强子成分上只结合了一个转录因子(缺乏另一个转录因子)，都不能形成能产型复合物(图 9-15a)。为什么一个转录因子与增强子元的结合不能形成促进转录的复合物呢？目前人们认为，尽管转录因子与增强子元的结合是各自独立的，但它们激活转录的功能却是互补的，谁也离不开它的伙伴。也就是说，结合于同一增强子成分的两个转录因子产生一个激活结构域(activation domain)，即图 9-15 中的∧∧。由于增强子成分是可以互换的，增强子元也可以互换，因此不同增强子成分产生的激活结构域在功能上一定是等价的。这也解释了为什么增强子成分必须是双部位的结构。

增强子成分又为什么需要两个以上才能起作用呢？这是因为能产型复合物并不直接与 RNA 聚合酶发生作用，而是与近启动子成分复合物发生相互作用。在近启动子成分复合物上有两个与激活结构域相互作用的位点，因此需要两个增强子成分及其复合物。这种近启动子成分复合物到底是什么？有可能是 TATA 框及其结合因子TFⅡD与RNA聚合酶Ⅱ以及其它辅助因子所构成的。这样的相互作用示于图 9-15b。

增强子所包含的两个水平上的二元组织方式有什么意义呢？其中有一个意义是清楚的，就是可以利用有限的反式作用因子提供更多的调控花式。例如，如果有两种反式作用因子 A 和B作用于一个水平的二元结构的话，那末一共只有四种花式(AA，AB，BA，BB)。然而，仍

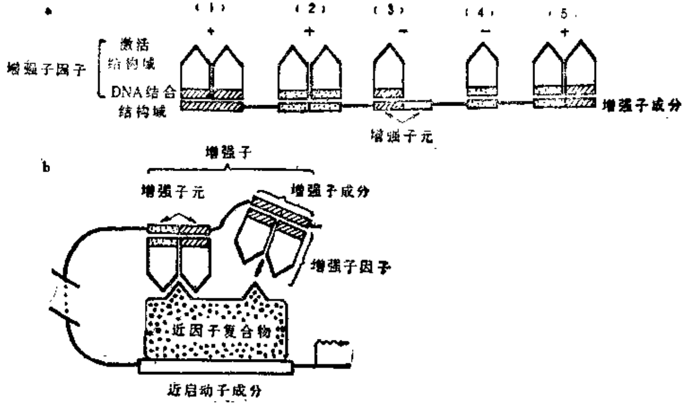



图 9-15 (a) 增强子元与反式作用因子的结合生成能产型复合物(1，2，5)和非能产型复合物(3，4) (b) 两个增强子成分复合物与近启动子成分复合物相互作用以促进转录的起始。其中双向箭头表示转录起始前后该处出现缔合-解离或解离-缔合物的交替过程，浪线箭头表示 RNA 合成的起始。

旧是 A 和 B 两种反式作用因子，如果作用于两个水平的二元结构，则可形成 16 种花式 (AA-AA，AA-AB，AA-BA……)。实际上，反式作用因子的种类还要多得多，产生的调控花式更是变化多端。这正是增强子的组织特异性的根源。只有当某种组织或细胞具备增强子所需的全部反式作用因子(1～4 种)时，该增强子才显示促进转录的功能。如图 9-15a3 中，该增强子成分上缺少一个反式作用因子，因而没有功能；如果在具有所需的两种反式作用因子的组织中，该增强子成分就会有功能。另一种可能的意义是：由于增强子的作用可以在不同的方向和很大的距离范围内促进基因的表达，如果其结构像启动于成分那样简单的话，很多序列都有可能因自发突变而成为增强于，从而造成基因表达的混乱。增强子的复杂的组织方式就阻止了这种情况的产生。

整个真核生物基因的顺式调控成分的组织方式示于图 9-16。

尽管真核生物增强子在组织方式上和作用特点方面与启动子成分有所不同，但是二者均是与反式作用因子相结合，然后通过反式作用因子的某种相互作用，从而促进了转录。在天然存在的增强子和启动子中，有些 DNA 序列及其反式作用因子是共同的，而且有些增强子元和启动子成分可以互换。例如免疫球蛋白基因的增强子中的八核苷酸序列 (ATTTGCAT) 也存在于许多基因的启动子中，当然它们都是与同一种蛋白质因子(NF-A)相结合。当 SV40 的增强子连接到 $\beta$-珠蛋白基因启动子的上游 100 bp 以上时，如果缺失两 UPE，则基本上丧失转录活性。如果以 SV40 的增强子代替两个 UPE，则很大一部分转录活性可以恢复，由此看来，启动子与增强子有很大的相似性。有趣的是，热震惊基因的上游启动子成分 HSTE (CNNG-

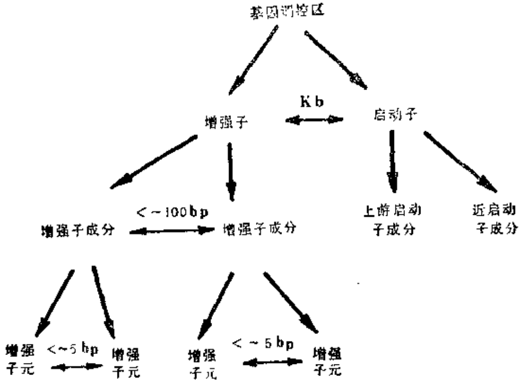



图 9-16　真核生物基因的顺式调控成分的组织方式示意图

AANNTTCNNG)如果放在其他启动子的远上游,则没有什么作用;如果重复放置两个这样的序列,则变为典型的增强子。然而,并不是所有的上游启动子成分都可以转变为增强子的,例如 CAAT 框的多拷贝重复就不能在远距离刺激转录。

增强子为什么能远距离地刺激基因的转录?近年来的许多资料都说明,结合于增强子上的蛋白质因子与启动子上的蛋白质相互作用,形成复杂的转录复合体。在此过程中,两个结合位点之间的 DNA 形成噜噗结构(图 9-15b)。Atchison 等人通过对免疫球蛋白k链的启动子和增强子关系的研究,在k增强子存在时,距增强子 440bp 与 2.7kb 的两个启动子具有同样的活性,从而否定了所谓增强子是 RNA 聚合酶双向进入位点的假说。按照双向进入位点的假说,就必然是离增强子近的启动子的转录效率高。

近年来人们还发现一种与增强子作用相反的顺式调控成分,叫做抑制子(Silencer)抑制子的作用机制还不清楚。人们推测,抑制子的作用机制可能与增强子类似,只是效应相反。

### 三、基因调控的反式作用因子

目前已发现了许多反式作用因子,其中主要是蛋白质。其中研究得最多的就要算识别 GGGCGG 的转录因子 SP1,识别 CCAAT 框的 CTF,热震惊基因的转录因子 HSTF,固醇类激素受体等。最近,人们发现同形异位基因编码的蛋白质也是转录因子;有些转录因子,如 Ap1竟是由原癌基因*jun*所编码的。这些都引起了人们极大的兴趣,它向人们提供了胚胎分化和癌症发生的新线索。下面,我们仅以SP1和CTF为例,让读者了解转录因子的一般情况。

**1. SP1**　是从人类细胞中分离到的一种转录因子,是SV 40 转录所必需的。它所识别的序列 GGGCGG 是非对称的,而且其作用是双方向的。在 SV40 的启动子中,有 6 个这样的 GC 框,SP1 与其中三个结合较强,两个结合较弱,还有一个不结合(图 9-17)。每个 SP1 可以保持大约 20bp 的 DNA。SV40 启动子中的 GC 框可促进早期和晚期基因的转录。现在已经发现许多核基因的启动子都含有GC框,如缀金蛋白(metallothionein)基因,ADP核糖基转移酶基因等。人们还发现艾滋病毒基因的启动子中也含有 GC 框。不同的启动子中GC 框具有不同

的排列方式。图 9-17 表示 SV40 早期和晚期基因的启动子，疱疹病毒 IE3 基因启动子，疱疹病毒胸腺嘧啶核苷激酶基因启动子，猴基因组中双方向启动子（β- 结合区域），小鼠二氢叶酸还原酶基因启动子等启动子中 GC 框的位置和方向（以富含 G 的一条链的方向作为 GC 框的方向）。根据图中 19 个 SP1 结合位点的序列进行分析，得出 GC 框的一致序列是

$$ {}^{G}_{T}GGGCGG^{GGC}_{AAT} $$

SP1 均结合于 GC 框的大沟中，而且各 SP1 均结合于 DNA 双螺旋的同一侧。SP1 不具备组织特异性，启动子的 GC 框上 SP1 的结合与否可使转录效率相差 10~25 倍。

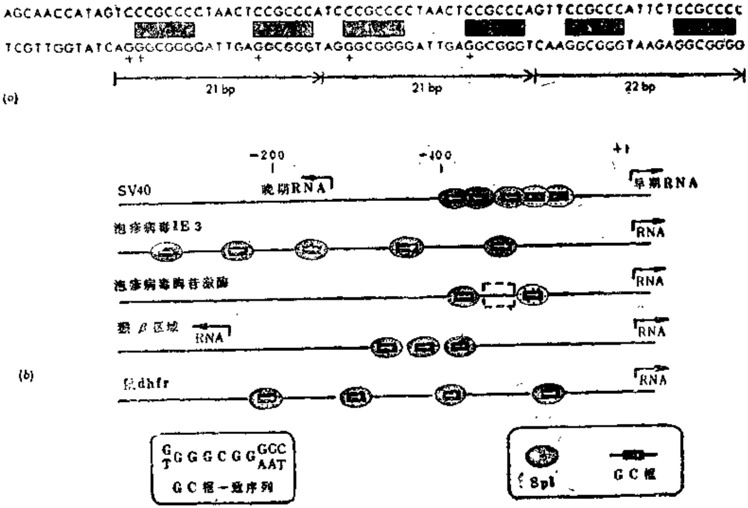



图 9-17　转录因子 SP1 及其结合位点 GC 框

(a) SV40 启子动中的 6 个 GC 框

(b) 不同的启动子中，GC 框具有不同的位置和方向

2. CTF　是真核生物中结合CCAAT框的一个转录因子，CCAAT框的存在较为普遍。除了在小鼠 β-珠蛋白启动子中CCAAT框需要与另一个启动子成分 CCACACCCG 相互作用以外（实际上为两种结合的蛋白质因子相互作用），其余的CCAAT框均独立地起作用。体外转录试验表明，在 HeLa 细胞核提取液中，有一种蛋白质因子能够结合疱疹病毒胸腺嘧啶核苷激酶启动子的 CCAAT 框以及小鼠的 α- 珠蛋白基因的 CCAAT 框。

通过对大量的启动子结合试验，发现 CTF 的识别序列与过去发现的能够识别腺病毒复制原点的核因子 1 (NF-1) 的识别序列很相似。通过遗传学方法和生物化学方法，都证明 NF-1 是腺病毒 DNA 复制起始绝对不可缺少的。通过与大量的不同病毒和细胞基因组 DNA 亲和性试验和单碱基替代对于 NF-1 亲和性的影响的试验，得出 NF-1 的一致识别序列为 TGG(A/C)$N_5$GCCAA。这一一致序列与CCAAT是相关的，只不过NF-1的一致序列具有二倍旋转对称性，而CCAAT缺少这种对称性（图 9-18）。对这两种蛋白质因子进行进一步的生

化分析，发现它们完全是同一种分子。同一种DNA结合蛋白质参与了完全不同的细胞生理功能，这一发现是耐人寻味的。CTF(NF-1)是如何促进复制过程的，现在还不清楚，或许与第一个dCMP结合有关(见第三章)。近来人们发现，能与CCAAT框结合的并不止NF-1，还有NF-Y等。

CTF (NF-1)与SP1一样，同样是作用于DNA的大沟，也没有组织专一性。迄今人们尚未发现能够提高转录活性达100倍以上的组织专一性因子，而实际上真核基因在不同的组织中表达可能相差好几个数量级，如鸡的卵清蛋白基因。这就表明，基因转录的激活可能需要多种蛋白质因子的共同作用。

| 序列 | 来源 | 结合力 |
|---|---|---|
| TTGGCT NNN AGCCAA | 一致序列 | |
| TTGGAT TGA AGCCAA TA | Ad-2/5 复制原点 | 高 |
| ATGGCG CGC AGCCAA TG | ras (-300) | 高 |
| ATGGTA TGG CGCCAA GA | 人 - 珠蛋白 (-200) | 高 |
| CTCCGC GCC AGCCAA TG | 人 - 珠蛋白 (-70) | 高 |
| TCCCTT CTG AGCCAA TC | hsp70 (-70) | 中 |
| CCTTCC TTG GACCAA TA | hsp70 (-145) | 中 |
| AAGC CAG AGCCAA CC | 人β - 珠蛋白 (-150) | 中 |
| GGGCCG CTC AGCCAA CC | ras (-250) | 中 |
| CGGGGA CAC CGCCAG CA | HSV tk (-160) | 中 |
| GTTCGA ATT CGGCAA TG | HSV tk (-80) | 低 |
| CTGAGG GTT GGCCAA TC | 人β - 珠蛋白 (-20) | 低 |

图9-18 CTF (NF-1)的各种结合位点的比较结合力的高、中、低依次相差5倍以上一个因子，是按照DNAaseI脚印法分析而估计的

## 四、真核基因转录调控的机制

**1.** *顺式调控成分与反式调控因子的相互作用* 基因的所有顺式调控成分，包括上游启动子成分(UPE)和增强子，都要与相应的反式作用因子结合。反式作用因子与顺式调控成分结合以后，还要通过蛋白质与蛋白质之间的相互作用(包括反式作用因子之间的相互作用和反式作用因子与RNA聚合酶之间的相互作用)，才能实现它们对子基因转录的调控。关于调控蛋白质与其识别位点之间作用的分子机制以及蛋白质分子之间相互作用的可能方式将在下一章中讲授。

**2.** *可诱导的顺式调控成分* 可诱导的顺式调控成分主要是指那些对热震惊、重金属、病毒感染、生长因子、固醇类激素等能作出反应的基因调控成分。这些成分中有的与基因相距很远(增强子)，有的就成为UPE成分。

不同物种的同一种可诱导基因的顺式调控成分有很大的保守性，如各种哺乳动物的缀金蛋白启动子在－50，－80和－140处各有一个富含GC对的12 bp的保守序列(TGCGCCCG-

GCCC)。各个基因被重金属(如 $Cd^{++}$) 所诱导的水平常取决这些成分的拷贝数。这在可诱导的顺式调节成分中是一个普遍的现象。例如在果蝇的 *hsp* 70 基因 TATA 框上游方向 25 bp, 50 bp, 150 bp, 220 bp 附近,共有四个 HSE 成分。然而,在不同的组织中达到有效诱导所需的重复成分的数目不同。这可能反映了不同组织中,反式调节因子种类的不同。

顺式调控成分对基因转录的激活作用可以通过不同的途径(图 9-19),可能是一种正调控因子的激活(或合成?);也可能是使阻遏蛋白失活;也可能是上面两种作用同时发生。果蝇 *hsp* 70 基因的 HSTF 与 HSE 的作用就是正调控的例子。尽管在诱导和非诱导的细胞中都能分离到 HSTF,但在诱导细胞中分离出来的 HSTF 具有高得多的体外转录活性。这就表明,HSTF 不是诱导合成的,而是在诱导之后HSTF受到了修饰或者结合了小分子配基。在迄今所发现的反式调控因子中,没有一个是诱导时合成的。也就是说结构基因已经作好了预备状态,随时可以被诱导。HSTF 是以二聚体形式结合于 HSE 的。第一个二聚体的结合引起了 DNA 的弯曲,从而促进第二个二聚体的结合。第二个二聚体的结合引起了 HSTF 和 DNA 构象变化,从而触发了转录反应。

高等真核生物中研究得最清楚的反式激活因子就是许多固醇类激素的受体,它们能激活固醇类激素所诱导的基因。目前已出现了多种固醇类激素和它的受体以及激素-受体复合物在 DNA 上的结合位点(顺式调控成分)。雌激素、黄体酮、糖皮质激素等激素受体的基因均已被克隆并进行了序列分析。一个质粒携带受体蛋白基因,另一个质粒含有可为相应激素所诱导的增强子。将这两个质粒共同转化到不含受体的细胞中,激素受体复合物与增强子的结合促进了有关基因的转录。激素调节转录的组织特异性就在于靶细胞含有大量的激素受体(如鸡输卵管组织含有雌激素受体),而非靶组织不含受体或受体很少。早期研究表明,固醇类激素与细胞膜上的受体结合以后,复合物进入细胞核,再结合于DNA 上。后来发现,在有无雌激素存在时,细胞核内都有其受体。人们还发现通过用免疫亲和层析纯化的糖皮质激素受体,在没有激素时在体外能与其调控位点结合,而用一体外合成的受体在没有糖皮质激素时却不能与其调控位点结合。因此,激素对于基因调控的机制可能不是那么简单。然而,激素对基因调控作用的轮廓是清楚的,激素—受体—顺式作用位点,三者缺一不可。

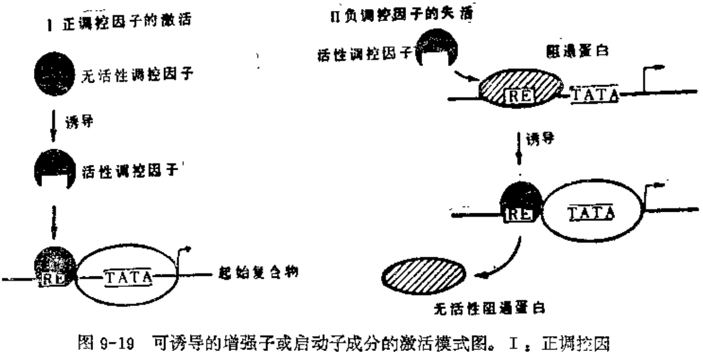



图 9-19 可诱导的增强子或启动子成分的激活模式图。I:正调控因子的激活;II,负调控因子的失活。诱导同时导致正调控因子的激活和负调控因子的失活的机制没有在图中表示出来

以上的几个例子都是正调控因子的激活。具有正调控因子的激活和负调控因子失活的例子是 β- 干扰素基因的调控。β-干扰素基因启动子调节成分(IRE)有两个功能结构域所组成。即一个负调控成分和一个相邻或少许重叠的正调控成分。当负调控成分缺失时，引起基因在无诱导物时高水平转录(由此发现 β-IFN 基因负调控成分的存在)。后来人们发现，将这种负调控成分与异源启动子组合，则阻遏与之相连的基因的转录；若代之以正调控成分，则出现增强子效应。用脚印法(用 DNAase Ⅰ)发现，负调控成分在诱导前受保护而在诱导后不受保护，而正调控成分的情况恰好相反。很可能正调控因子激活后结合于正调控位点而同时排斥了负调控因子的存在。

正调控因子的激活和负调控因子的失活的作用模型示于图 9-19。

**3.** 组织特异性的顺式调控成分　基因表达的高度组织特异性的一个例子是免疫球蛋白(Ig)基因的表达不仅限于 B 细胞，而且必须在类淋巴细胞分化的特定阶段才能被激活。Ig 基因表达需要三个 DNA 序列：B 细胞特异性的增强子；Ig 启动子以及细胞专一性的转录后控制所必需的基因内序列。通过增强子和启动子之间的协同作用有效地控制着 Ig 基因的表达。这种协同作用可能是通过结合于两种 DNA 成分上的蛋白质因子而实现的。在增强子和启动中都有 Ig 八核苷酸序列 (ATTTGCAT)，它们结合了相同的蛋白质因子 NF-A1 和 NF-A3。这两个因子是非组织特异性的，因此 Ig 基因表达的组织特异性在于其增强子上的增强子核心序列的作用。

B 细胞特异性的增强子存在于 mu 重链和 kappa 轻链基因的内元中。通过缺失突变的研究，表明 mu 重链基因的增强子至少包含 140bp。其中除了一个上述的 Ig 八核苷酸外，还有四个与 CAGGTGGC 相关的核心序列，它们结合不同的蛋白质因子(图 9-20)。但是体外结合试验只发现 $E_1$，$E_3$，和 O (八核苷酸序列)结合了蛋白质因子。其原因还不清楚，估计可能是与之相应的蛋白质因子失活了。五个作用位点中任何一个缺失，只会降低而不会完全消除增强子对转录的激活作用。

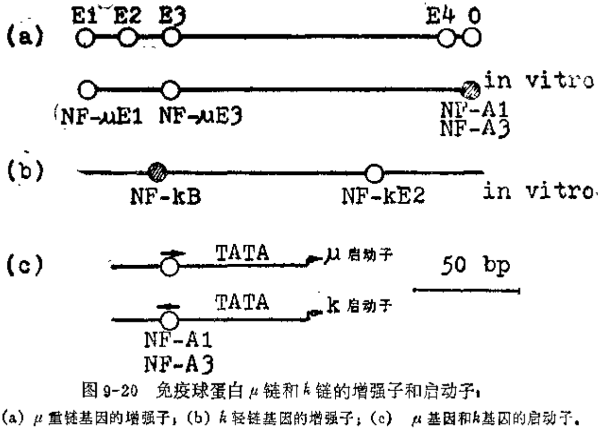



图 9-20　免疫球蛋白 μ 链和 k 链的增强子和启动子：
(a) μ 重链基因的增强子；(b) k 轻链基因的增强子；(c)　μ 基因和 k 基因的启动子。

有趣的是，$E_1$，$E_3$，O 与蛋白质因子的体外结合并不限于用 B 细胞的提取液；然而，基因组脚印试验表明，只有在 B 细胞中这些结合位点才得到保护。人们的解释是，这些结合因子是

各种细胞中广泛存在的，但是这些位点只有 B 细胞中才能被结合。这就表明这些因子在 B 细胞内必须被修饰成活性形式才能结合；而体外试验可能是由于剂量大而引起结合。在 B 细胞中，存在两种能结合kappa 链基因增强子的蛋白质因子，NF-κB 和 NF-κE2。近来发现，NF-κB并不仅仅存在于B细胞中，在许多其他细胞中都存在着非活性形式的 NF-κB。在 B 细胞的前体细胞中也无可觉察的活性形式 NF-κB。但是当 B 细胞用脂多糖(LPS)，phorbol 酯类，环己亚胺等试剂处理时，则可以在核提取液中测得NF-κB的活性。上述试剂都是已知能诱导 kappa 基因表达的试剂。使人吃惊的是，不能够表达 kappa 基因的 T 细胞和 HeLa 细胞经上述试剂处理也能诱导出有活性的 NF-κB。人们已经知道上述试剂能促进细胞内蛋白质激酶 C 的产生。因此 NF-κB 的修饰有可能是采取磷酸化的形式。

调控基因转录的调控蛋白质基因的表达又是如何调控的？如果还是采取上述的调控机制的话，就必须有无穷尽的连瀑式调控体系。这显然是不必要的，也是不可能的。由于调控蛋白质需要量极少。在细胞内可能是低水平的本底组成型表达，并有可能产生反馈抑制。这些基因产物最初是以无活性形式存在于细胞内。引起细胞分化诱导因素可能直接地通过翻译后修饰迅速地将无活性的调控蛋白质转化为活性形式，从而促进有组织特异性的基因的表达。

## 第五节　rRNA 基因的转录：先终止还是先起始[66~74]

真核生物的核糖体 RNA 基因先转录成一个较长的(随着物种的不同，可以 35S 到 47S，即 6 至 15 kb)前体分子，然后才进行转录后处理成为成熟的 18 S，5.8 S，28 S rRNA，以便参与核糖体的组建。尽管真核生物细胞中常有 10 000 种以上 RNA 分子，但其中大约一半数量是 rRNA。因此 rRNA 的合成不但要与常规代谢相协调，而且要与周围环境、生物体发育阶段等条件相协调，以使细胞中有适当数量的核糖体。这就需要增加基因“剂量”，并需要极为有效而又高度控制的转录。在真核细胞中，随物种的不同，有 100～5 000 rDNA 拷贝串联成为重复序列，分布在少数的染色体区域，个别情况下以染色体外分子形式存在。这些 rRNA 基因，是由专一性的 RNA 聚合酶 I 进行转录的。一般情况下，这些 rRNA 基因拷贝数是不变化的，因此调控主要依赖于转录起始的速率。

对于大多数真核生物来说，rRNA 基因启动子可以分为两个部分：－40～＋5 称为近启动子，其功能决定转录起始的精确位置；－165～－40 称为远启动子，其功能是影响转录的频率。每种生物都有特定的转录因子与 RNA 聚合酶 I 结合，从而促进与启动子形成转录起始复合物，因此 RNA 聚合酶 I 具有明显的种族特异性。启动子均位于所谓“不转录的间隔区”(Nontranscribed Spacer，简称 NTS)。首先应该指出，NTS 是一种误称。现在发现大部分的 NTS 是转录的。非洲爪蟾的 NTS 结构如图 9-21 所示。

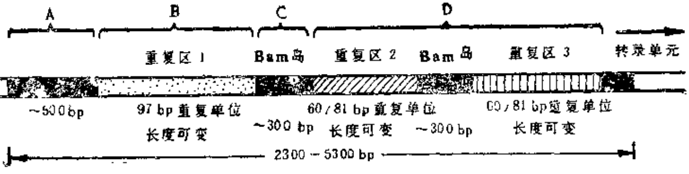



图 9-21　非洲爪蟾 NTS 的结构

在图 9-21 中，60/81 bp 与 Bam 岛重复 2～6 次，最多可重复 8 次。其中 A 区在序列和长度上都是独特的。而 Bam 岛(因含有 Bam H1 切点)也都是较短的不变序列。每个 Bam 岛的后半部与启动子及其附近序列十分相似。这表明，在 NTS 中可以由这些启动子发动转录。转录产物与前一个 rRNA 基因的转录产物一样，都终止于离下一个 rRNA 基因转录起点之前 215 bp 的 T3 序列。过去曾认为 T1(成熟的 28 S rRNA3′ 端)，T2(T1 下游 235bp)是转录的终点。现在已经知道，T 3 是唯一的转录终点。从 −200到 −194 的 7 bp 保守序列(GACTTGC)是转录终止子的不可缺少的成分。人们还发现，这一序列的突变，不但妨碍转录的终止，产生通读产物；而且还降低下一基因的转录效率。在小鼠 rRNA 的 −168 bp 处，也存在着一个转录终止子 To，To 的缺少也会产生类似的结果。T3 与其下游的启动子之间的距离也是关键的，插入 4bp 显著减少下一个基因的转录；如果插入 29bp 与缺失整个 T3 序列的影响是一样的。

细胞内有许多因素能促进 rRNA 基因的有效转录。专一性的 RNA 聚合酶 I 的使用就避免了别的基因启动子的竞争；rDNA 串联重复基因组织于很小的核仁区域，就增加了启动子的局部浓度；串联重复基因本身就使得 RNA 聚合酶在转录前一个基因后并不解离而活化了第二个基因的转录；NTS 序列中的多拷贝启动子能够吸引更多的 RNA 聚合酶 I，使 RNA 聚合酶 I 在终止子处激活了下一个基因的转录。或者有细胞核因子促进转录的终止，或者是 RNA 聚合酶正好碰到了一个牢固结合于 DNA 的转录起始因子。这些都还没有见分晓。但是，原核生物 λ 阻遏蛋白在 $O_{R1}$ 和 $O_{R2}$ 的结合，一方面阻遏了 $P_R$，一方面却激活了$P_M$。*lac* I 的转录可以通过 *lac* 阻遏蛋白结合于其操作子而形成自我调控。这些机制无不可以借鉴。现在暂时撇开这些未知机制，NTS 内的终止子序列能够促进下一个基因的转录，这已是毫无疑义的了。这正是"终止在先，起始在后"(第一个基因除外)。rRNA 基因的组织方式以及专一的 RNA 聚合酶 I，使rRNA基因的转录能够采取串级放大的机制，而这种机制是处于严格控制之下的。

## 第六节　意外的意外：RNA 聚合酶 Ⅲ 启动子的多形性和三种RNA聚合酶及其启动子的共通性[56,75～79]

在寻找真核生物 RNA 聚合酶Ⅲ的启动子特征时，许多研究者都相信它一定位于转录起始点的上游，就像原核生物基因的启动子一样。然而意外终于发生：1980 年 Sakonju 等人发现，非洲爪蟾 5S RNA 基因的启动子竟在基因的内部(见第六章)。在 RNA 聚合酶Ⅲ的基因内启动子这一事实已在教科书中确立其地位之后，人们又发现某些Ⅲ类基因的启动子除了基因内成分外，还有些基因外成分也有作用。特别是在应激反应中，转录起始点上游 25 bp 内的序列对转录有显著的影响。然而，基因内成分仍然是主要的，上游序列对转录的影响并不使人感到意外。

在 RNA 聚合酶Ⅲ启动子问题上的第二次意外又发生了，1987 年 Murphy 等人发现，由 RNA 聚合酶 Ⅲ 转录的基因 U 6 和 7 SK 在它们的转录起始点下游序列全部缺失，仍然能够进行活跃的转录。也就是说，这两个由 RNA 聚合酶Ⅲ转录的基因的启动子全部位于基因的上游，是外部启动子而不是内部启动子。尽管这两个基因确实含有 RNA 聚合酶Ⅲ启动子的共有的基因内成分 A 区序列，但 A 区序列并不是这两个基因转录所必需的。这样，RNA 聚合酶Ⅲ的启动子在不同基因中可以位于 −3 的上游，也可以位于 +50 的下游。过去人们曾期待同

种 RNA 聚合酶的各种启动子一定处于大体相同的位置。

许多不同的 RNA 聚合酶Ⅲ转录的基因，在其转录起点上游都有较保守的序列，特别是在 -25 附近有一个 TATA 样序列。在基因 7 SK 的 TATA 样序列中的点突变都强烈地降低它作为 RNA 聚合酶Ⅲ转录模板的性能。还有许多基因也是如此。在某些基因中，转录起始点附近的序列也有明显的重要性。TATA 样序列是否结合有在Ⅱ类启动子中起作用的 TATA结合因子，还没有确凿的证据，但看来是很有可能的。这就是说，真核生物的不同 RNA 聚合酶可能使用共通的转录因子。其实这也并不奇怪，近年来人们已发现真核生物的三种 RNA聚合酶也有些共通的亚基。

序列分析已显示出由 RNA 聚合酶Ⅱ转录的 U1 和 U2 基因以及由 RNA 聚合酶Ⅲ转录的 U6 基因在 -70 附近都有同样的上游启动子成分。值得注意的是，如果 U6 基因缺失这样一个上游启动子成分，则不再能被 RNA 聚合酶Ⅲ所转录。由此可见，U 系列 RNA 基因的转录是由一个共同的转录因子所介导，而不管是由聚合酶Ⅱ还是聚合酶Ⅲ所转录。

对于 U 6 和 7 SK 基因转录的研究还得出另一个令人惊讶的结果：这两个基因的远上游序列（-200 bp 以上）对于 RNA 聚合酶Ⅲ的转录有 10～20 倍的刺激效应。在这段序列中，有两个拷贝的八核苷酸序列花式（*ATTTGCAT*）。八核苷酸序列存在于许多Ⅱ类基因的增强子中（如免疫球蛋白基因）和上游启动子成分中（如 U1 和 U2 基因）。实验表明，RNA 聚合酶Ⅲ对 U6 基因的转录能够被外加的 U2 基因竞争性地抑制，而不能被八核苷酸序列处突变了的 U2 基因所抑制。脚印试验表明，同一种蛋白质因子 NF-A 能够结合于聚合酶Ⅲ所转录的 U6 基因的八核苷酸序列和聚合酶Ⅱ所转转录的 U1、U4，免疫球蛋白基因的八核苷酸序列。在 U6 基因中的两拷贝八核苷酸序列的作用很像增强子，尽管它有点位置和方向依赖性。现已表明，增强子不再是Ⅱ类基因所特有，而是能增强所有三种 RNA 聚合酶的转录。

三种 RNA 聚合酶转录的基因具有共同的特征和共同的转录因子，这其中暗示着某种调控机制。例如 U4 和 U6 基因产物是按 1:1 比例结合成复合物，但分别由聚合酶Ⅱ和聚合酶Ⅲ所转录，二者之间可能就是通过这种方式进行协调的。这两类基因间协调的桥梁可能就是 TATA 结合因子和八核苷酸序列结合因子。由 RNA 聚合酶Ⅰ转录产生的主体 RNA（18 S, 5.8S, 28 S）和由 RNA 聚合酶Ⅲ转录的 5S RNA 也是按 1:1:1:1 比例参加核糖体构建的，这两类基因的启动子有何共同的序列特征和共同的转录因子还有待进一步研究。至于有些基因，如 *c-myc* 基因在离体试验中既可被 RNA 聚合酶Ⅱ转录，又可被聚合酶Ⅲ所转录。在体内的情况如何及其可能的调控意义都还不清楚。

## 第七节　真核生物中转录的翻译控制[61,64,80～83]

在原核生物中，没有核膜结构。当一个基因刚开始转录不久，核糖体就会结合到正在转录的 mRNA 上，造成转录与翻译的耦联。这正是原核生物中衰减子调控机制的结构基础。而真核生物具有核膜结构，人们曾经否定了翻译影响真核生物基因转录的可能性。然而，近年来的研究表明，真核生物的基因转录也存在翻译控制机制。

在啤酒酵母中，不同的氨基酸合成途径处在一个通用机制的控制之下。例如缺少色氨酸时，能够解除色氨酸，精氨酸，组氨酸，异亮氨酸，赖氨酸等氨基酸合成酶系的阻遏。这一交叉调控途径就称为通用氨基酸控制，这一控制机制至少影响到 24 个基因的表达。

通用控制发生在转录水平上，因为上述任何一种氨基酸饥饿都造成这些基因 mRNA 的增加。

如果这些互不相连的基因中，每个基因的上游都存在着一个顺式作用位点，则有可能通过一个共同的反式作用因子实现这种控制。事实上，人们已经发现这些基因的 5′ 端非编码区域含有一小段重复的 TGACTC，这是通用控制机制所不可缺少的。如果将这样的控制序列与无关的基因(如细胞色素 C 基因)重组在一起，则使该基因也处于通用氨基酸控制之下。现在人们已经发现了这一反式作用因子是基因 *GCN*4 的产物。在体外实验中，*GCN* 4 蛋白能强烈地结合于顺式作用序列 TGACTC。

上述控制机制正是我们在第四节中所讲的机制。如果只是如此，那还谈不上是翻译控制。而 *GCN*4 基因的表达是处于翻译控制之下。*GCN* 4 mRNA 有一个～600 核苷酸的非编码的前导序列。这一序列含有四个很短的开放阅读框架，每一个都含有一个起始密码子和一个终止密码子(图 9-22)。酵母 mRNA 一般只含有 30～90 个核苷酸的前导序列，而且一般不含同框的起始密码子和终止密码子。可见这 600 个核苷酸前导序列非同寻常。将 *GCN*4 基因与 *lacZ* 融合，则融合基因的表达依赖于另外几种不相连的反式作用基因 (*GCN*2 和 *GCD*1)。如果将四个短开放阅读框缺失，则 GCN4-Lac*Z* 融合蛋白增加，而没有相应的融合基因 mRNA 的增加。这就表明，四个开放阅读框是在翻译水平上控制 *GCN*4 基因的表达。此外，在含有四个开放阅读框缺失突变的菌株中，*GCN*4 基因的有效表达就不再依赖于 *GCN*2 基因产物。人们的进一步研究发现，这四个开放阅读框具有不完全相同的作用机制。第三个和第四个开放阅读框的抑制作用可被第一个和第二个开放阅读框所解除。第三和第四个开放阅读框又处于两个反式作用因子的控制之下：GCD1 产物是它们的抑制效应所必须，而 GCN2 产物用以克服其抑制效应。正是这两个反式作用因子作用于四个开放阅读框，调节了 *GCN*4 基因对氨基酸水平的应答反应。而 GCN2 和 GCD1 这两个基因产物的性质和作用机制仍然有待于进一步的研究。

以上分析表明，酵母基因的转录是受到翻译调控的。但是这种机制不同于细菌中的直接耦联机制，而是要分为两步：首先是共同的反式作用因子本身的翻译调控，然后这一因子透过核膜，实现了对一组处于通用控制之下的基因的调控。

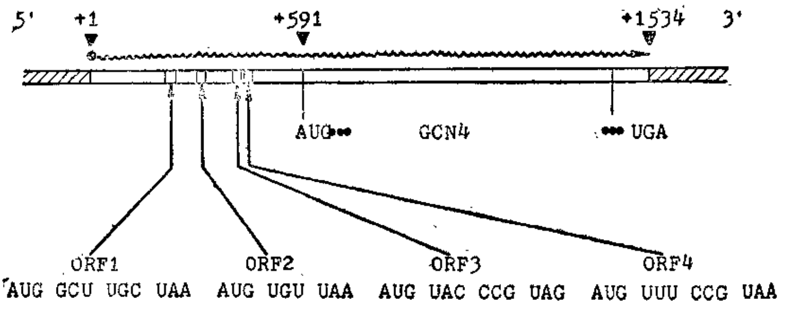



图 9-22 基因 *GCN*4 的前导序列中有四个短的开放阅读框

## 第八节 转录后水平的调控[80,84~89]

### 一、hnRNA 的选择性加工运输

真核生物基因活性在转录水平调控是决定细胞质中 mRNA 水平的一个重要(在大多数情况下是主要)方式。可是在试验中人们观察到，某些基因序列在 hnRNA 和 mRNA 中的相对丰富度差异很大。换言之，一些基因的转录活性和它们在细胞质中 mRNA 拷贝的数目不成比例；另外，虽然生长中的细胞和静止细胞的hnRNA 转录速度相差不多，含多聚A 的 hnRNA 比例差异也很小，但是生长中细胞的细胞质中 mRNA 含量和种类比静止细胞要多得多，这些结果说明除了转录水平的调控外，hnRNA 的加工及 mRNA 的运输也是决定细胞质中各种 mRNA 分子水平的调控环节。细胞可以在不同情况下有选择性地将不同的 hnRNA 加工成 mRNA。并把它们运输进入细胞质。

hnRNA 选择性加工运输对于基因调控的重要性在海胆 *S. purpuratus* 中表现得非常明显。在胚胎时期的囊胚细胞中大约有 20 000 种不同序列的 hnRNA，其中约有 13 000 种加工形成 mRNA，在成体肠细胞中，大约有 25 000 种 hnRNA，可是只有大约 3 000 种形成 mRNA。在尿胚细胞中存在而在肠细胞中没有的mRNA序列极大多数可以在肠细胞 hnRNA 中发现。这表明，在海胆中许多基因的转录并不因为组织的不同而有很大差异，不同组织调控自己 mRNA 水平的主要方式似乎不在转录水平而在对 hnRNA 的选择加工。似乎只有一小部分基因的调控发生在转录水平。

目前对 hnRNA 选择性加工运输的机制了解得很少。hnRNA 的加工涉及到 5′ 帽子的形成，3′ 端多聚 A 的添加，RNA 拼接等许多步骤，调控作用可能发生在其中的每一步。加工完成后，mRNA 能否进入细胞质还要看核酸酶的降解情况以及穿越细胞核膜的能力等。在以上这些步骤中，hnRNA 不是以游离状态存在的，而是自始至终和蛋白质形成核糖核蛋白(hnRNP)颗粒，很可能 RNP 颗粒的形成是 RNA 加工的先决条件。RNP 复合物中的蛋白质组分可能包括了加工和运输各步骤需要的酶系统和接受调控信号(如激素)的受体蛋白。此外，RNA拼接(可能还有运输)过程需要 snRNA 与蛋白质形成的复合物 snRNP。

mRNA 在转录开始后不久就有核蛋白与之结合而形成 hnRNP 颗粒。这种颗粒排列于 mRNA 分子上，呈串珠状，就像核小体结构一样。在电子显微镜下，一个正在转录的活跃基因周围放射出许多串珠样结构(图 9-23)。然而 hnRNP 颗粒与核小体不同，其颗粒大而不均一(直径大约 200Å)，而且对 RNAase 敏感。在用核糖核酸酶进行温和的降解之后，可以得到沉降常数为 30 S~40 S 的颗粒，与核糖体小亚基差不多大小。然而 hnRNP 颗粒与核糖体不同，hnRNP 富含蛋白质，其中蛋白质占四分之三，RNA 占四分之一。这样，每个 30 S 颗粒中包含大约 600 个核苷酸的前体 mRNA 片断。RNA 在颗粒表面绕了几圈。hnRNP 中的蛋白质是一组高度保守的分子量为 30~40 KDa 的蛋白质。它们含有异常修饰的氨基酸(如二甲基精氨酸)，它们的抗原性亦高度相关，一种蛋白质的单克隆抗体常与这一组中的其他蛋白质有交叉反应。在细胞核中，hnRNP 蛋白质也是含量较丰富的非组蛋白。现在还不清楚，为什么这组蛋白质只与前体 mRNA 结合而不与 rRNA 和 RNA 聚合酶Ⅲ转录产物相结合，这可能与 RNA 聚合酶Ⅱ的某一亚基(已知的或迄今尚未发现的)有关。hnRNP 蛋白质的任务仅限于细胞核内。当成熟的 mRNA 被运送到细胞质时，hnRNP 蛋白质即被完全不同的另一组蛋白质所取

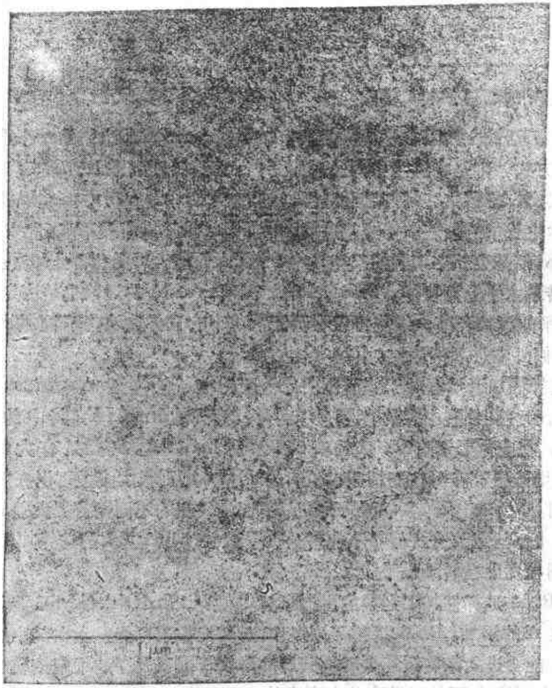



图 9-23 果蝇卵泡细胞中的活跃基因周围放射状的转录产物与hnRNP蛋白质形成 hnRNP 颗粒

代，形成 mRNP 复合物。

## 二、mRNA 前体的选择性拼接

结构相关的，发育上受控制的，具有细胞类型专一性的蛋白质异形体(protein isoforms)的表达是真核细胞分化的重要特征之一。产生这种蛋白质多样性的分子机制，就目前所知，可以分作两个范畴：一是多基因的基因家族中的各成员在特定的细胞中，在一定的发育阶段和在一定的生理条件下选择性的表达；二是从同一基因产生若干种有部分相同结构的蛋白质。后者是我们在这一小节中讨论的主要对象。

从同一基因产生不同的蛋白质，主要发生在转录后水平上，即通过 mRNA 前体的选择性拼接而产生不同的成熟 mRNA，然后翻译成不同的蛋白质。除了转录后水平之外，还可以发生在转录水平上和翻译水平上，但是在真核生物中从同一基因上产生不同的的蛋白质，绝大多数是通过 mRNA 前体的选择性拼接而进行的。

高等真核生物的蛋白质基因大多数是不连续基因，原初转录产物中所包含的内元序列通过RNA 拼接而被除去。尽管有时外元数目众多(如肌红蛋白重链基因的外元多达 41 个)，但也能有条不紊地将外元连接成成熟的 mRNA。这种拼接方式称为组成型拼接 (constitutive

splicing)。这样，通过组成型拼接，一个基因只产生一种成熟的 mRNA，一般也只产生一个蛋白质产物。与这种精确的组成型拼接相反，真核生物中有些基因的原初转录物有几种不同的拼接方式，因而产生不同的成熟 mRNA，翻译产生不同的蛋白质。有时，由于转录时选择不同的启动子或者在转录物上选择不同的 PolyA 位点而使"原初"转录物就不相同。在这种情况下，两种前体 mRNA 具有不同的二级结构，从而影响到拼接过程。而对于原初转录物相同的情况下产生的不同拼接方式，其分子水平的机制几乎一无所知，可能除常规的拼接机构(拼接体，如第六章所述)之外，还有别的因子参与拼接过程。这里的实质是供体拼接点（即内元 5′ 拼接点）和受体拼接点（即内元 3′ 拼接点）的搭配问题。常规拼接机构以外的因子可能影响到这种搭配。

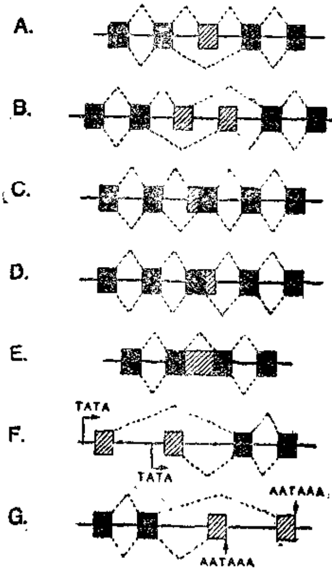



图 9-24 选择性拼接的类型。黑色方块表示组成型外元；斜线方块表示可选择的序列(外元或内元)；水平粗实线表示拼接过程中被去除的内元；各图上下的虚线表示两种不同的拼接途径。

(A) 一个匣子供选择；(B) 两个匣子供选择但互斥，(C) 两个 3′ 拼接点可供选择；(D) 两个 5′ 拼接点可供选择；(E) 内元保留；(F) 两个不同启动子；(G)两个 PolyA 位点

据目前所知，mRNA 的选择性拼接主要有图 9-24 所示的七种拼接类型。

(1) 匣子型：一个或几个外元可以出现在成熟的 mRNA 中，也可以在拼接过程中与内元一起被除去。这样，一个可选择的外元就会产生两种不同的成熟 mRNA，如果有 n 个可选择的外元，就会产生 $2^n$ 种成熟的 mRNA。这里的典型例子就是大鼠肌钙蛋白 (troponin) T 基因的外元 4，5，6，7，8。这五个外元的保留与否可产生 $2^5=32$ 种成熟 mRNA (图 9-25)。

(2) 互斥型：两个或多个相邻的外元分为两组，其一组在拼接过程中被保留在成熟的 mRNA 中，而另一组就必然被去除，反之亦然。这里，仍以大鼠肌钙蛋白 T 基因为例，其外元 16 和 17 在一个成熟 mRNA 中只会出现其中一个。这样，大鼠肌钙蛋白 T 基因利用上述两种类型的拼接方式就会产生 64 种不同的成熟 mRNA (图 9-25)。

(3) 内部拼接点型：包括内部受体拼接点和内部供体拼接点两种。这里所谓"内部"，通常是指外元的内部。但是，如果以另一种拼接途径为标准划分外元与内元，则可选择拼接点就可能跑到内元的内部了。内部受体拼接点型是一个内元的 5′ 拼接点可以与两个或两个以上的没有供体拼接点相隔的一组受体拼接点之一进行拼接。例如在大鼠的纤维结合蛋白 (fibronectin) mRNA 前体中有一个供体拼接点能与一组三个受体拼接点中的任何一个相拼接而产生三种成熟的 mRNA。内部供体拼接点型是一组没有受体拼接点相隔的供体拼接点中的任何一个都能与内元的 3′ 拼接点相拼接。内部拼接点型已在许多病毒中发现，如腺病毒，SV40 等。

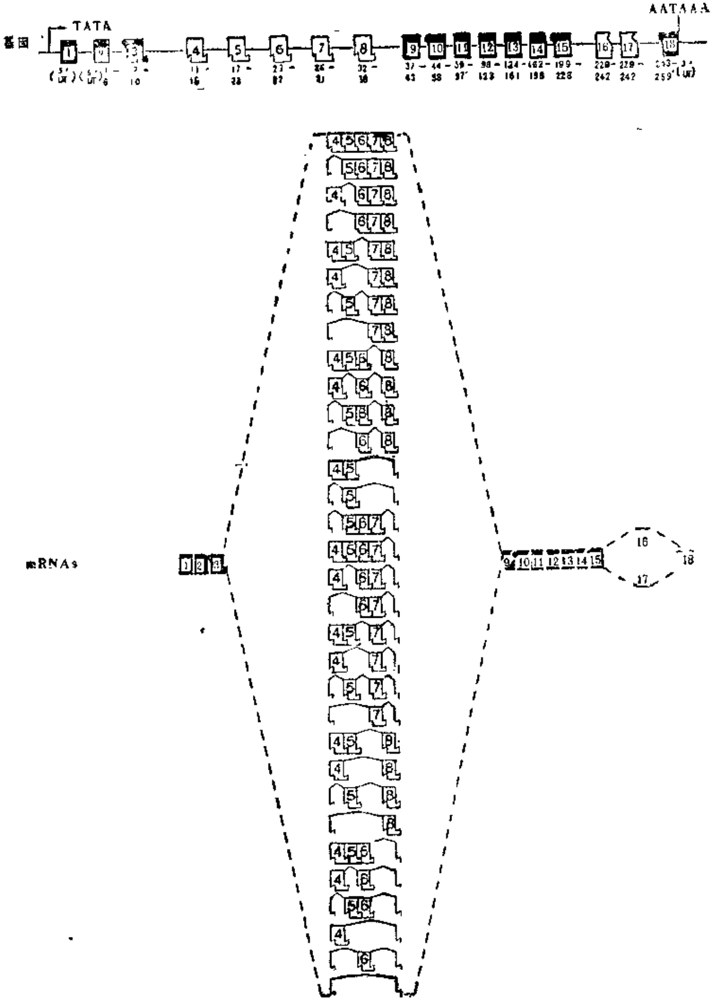



图 9-25 大鼠肌钙蛋白 T 基因转录产物通过匣子型和互斥型选择性拼接可以产生 64 种成熟 mRNA

(4) 内元滞留型：就是一个内元中的一对拼接点都没有参与拼接反应，因而将这一内元保留于成熟 mRNA 中。在大鼠的 γ- 纤维蛋白原基因的 mRNA 中，有 10% 保留了第七个内元的序列。在逆转录病毒中，从原病毒 DNA 上转录的 RNA 在拼接后作为 mRNA 使用，没有拼接的才能包装到病毒粒子中。

(5) 不同 5′ 末端型：鸡的肌红蛋白轻链在心脏中和砂胃中是不同的：在心脏中的亚基 ($MLC1_f$) 为 190 个氨基酸，在砂胃中的亚基 ($MLC3_f$) 为 149 个氨基酸。原因是这个基因具有两个启动子，二者相距 10 kb，因而造成两个原初转录物相差很大(一个 10kb，一个 20 kb)，长转录物经过拼接以后保留外元 1，4，5，6，7，8，9，其蛋白质产物为 190 个氨基酸；而短转录物经过拼接后保留外元 2，3，5，6，7，8，9，其蛋白质产物为 149 个氨基酸(图 9-26)。在长转录物拼接中，由于外元 2 缺少很好的受体拼接点，被排除在拼接产物之外是容易理解的，而对于外元 3 和 4 的取舍机制就不清楚了，正如所有的互斥型选择机制都还不清楚一样。这可能由于两种转录物大小相差甚多，形成的二级结构也不相同，从而影响到拼接模式。脊椎动物的肌红蛋白轻链基因结构大多如此。人的白细胞介素 2 的受体基因也有两个启动子，但相距很近，两种转录物有相互重叠的第一个外元。其第四个外元为匣子型选择性拼接，它是否与两种转录物的大小有关还不清楚，也许可能是独立进行的。

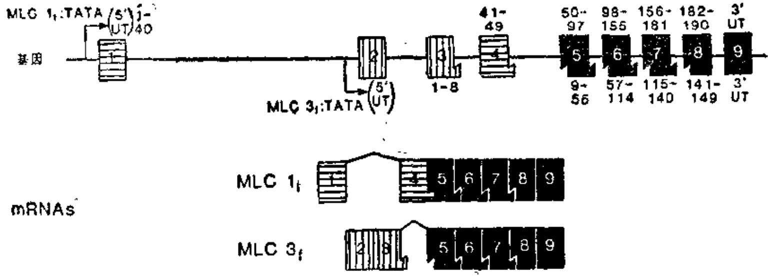



图 9-26 肌红蛋白轻链基因的组织形式和两种成熟 mRNA ($MLC1_f$ 和 $MLC3_f$) 的产生。黑色方块表示组成型外元；水平线条方块表示 $MLC1_f$ 所特有的外元；竖直线条方块表示 $MLC3_f$ 所特有的外元

(6) 不同 3′ 末端型：真核生物基因常有一个以上的 PolyA 位点，因此会在不同的细胞中产生具有不同 3′ 末端的m RNA 前体。具有不同3′末端的转录物会有不同的拼接方式。绝大多数真核生物基因的转录物是先加 PolyA 尾巴，然后再进行拼接。因此不同的细胞中会有不同的因子干预多聚腺苷酸化作用，最后影响拼接模式。但是，现在还不能排除细胞特异性因子先干预正在转录中的 mRNA 拼接，然后影响到 PolyA 位点的选择这一可能性。作为具有不同 3′ 末端转录物的例子是降钙素(calcitonin)相关基因的表达。在甲状腺 C 细胞中，转录物在第一个AAUAAA信号附近进行多聚腺苷酸化。这个短转录物内包含非编码外元，编码外元$C_1$ 和 $C_2$ 以及降钙素的编码外元(包括部分非编码区)。而在下丘脑中，转录物在第二个 AAUAAA 信号附近进行多聚腺苷酸化。这个长转录物除了包含短转录物中的由转录产生的全部序列之外，还包含了 CGRP 的编码外元和 CGRP 的 3′ 末端非编码外元。在长转录物的拼接过程中，外元 $C_2$ 的供体拼接点直接与 CGRP 编码外元的受体拼接点相连，从而删除了降钙素的编

码外元。两种成熟的 mRNA 分别翻译产生降钙素的前体和 CGRP 前体，然后通过酶促降解，产生降钙素和 CGRP 这两种激素。这一过程示于图 9-27。(注 CGRP 为 Calcitonin gene-related product 的缩写，意为降钙素基因相关产物，其生理功能还不太清楚)

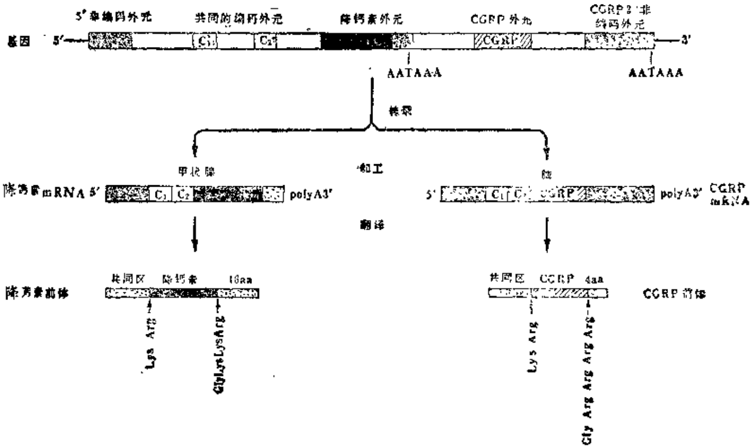



图 9-27　大鼠降钙素相关基因在不同的细胞中产生具有不同的 3′-末端的成熟 mRNA，先产生具有部分相同的蛋白质，再经酶促降解产生完全不同的两种激素。

## 第九节　翻译水平的调控[80,90~95]

翻译水平的调控对于真核细胞生命的重要性是很长时间以来有争议的问题。原核生物的基因大多组织成为操纵元的形式，转录出多基因的 mRNA。如果一个基因产物继续需要，而另一个不需要，那就只有通过翻译控制来解决。然而，真核生物基因的 mRNA 都是单基因结构，而且由转录到翻译，中间存在着许多控制步骤(如上节所述)。这样，真核细胞中翻译水平上的调控似乎是不必要的。然而，事实表明了真核细胞中确实存在着翻译水平的调控。这是一种迅速控制基因表达的方式，而且为各种高等真核生物所广泛使用。

### 一、mRNA 的稳定性

真核生物的 mRNA 稳定性，也就是 mRNA 的寿命，除了决定于内在因素(如 mRNA 的二级结构)外，还决定于转录后的修饰(戴帽子作用及帽子的种类，多聚腺苷酸化及 PolyA 尾巴的长短等)以及与一组细胞质蛋白质形成的 mRNP 中蛋白质组分的种类。此外，某些 mRNA 寿命还受到其他因素的影响，如哺乳动物乳腺细胞中酪蛋白 mRNA 的寿命可因促乳激素的存在而增加大约 20 倍。这里特别要强调的是，不同的 mRNA 会与细胞质中不同的蛋白质组分形成 mRNP，因而也具有不同的寿命。这里边的选择机制还毫无所知。原核生物 mRNA 的平均半寿期大约为 3 分钟，也就是说在一个细胞周期中，这些 mRNA 大约可以转化

10 秒。而快速生长的脊椎动物细胞，大约每天分裂一次，其 mRNA 的平均半寿期大约为 3 小时。

真核细胞中有许多编码“持家”蛋白质的 mRNA，这些 mRNA 一般是长寿命的，因为这些持家蛋白质在任何时候细胞都是需要的。然而，一些决定细胞周期的蛋白质(或许还有一些 RNA)，它们的合成必须按照严格的顺序开启或关闭。因此这些基因的转录产物必须是短寿命的。现在已知酵母有 50 多种与控制细胞分裂周期有关的基因(CDC 基因)。高等真核生物也必定存在着类似的 CDC 基因。

另一方面，在高等真核生物的最终高度分化的细胞中，许多 mRNA 极其稳定。网织红细胞就是一个很好的例子。在网织红细胞中已没有 RNA 合成，然而却在进行旺盛的翻译活动，合成的蛋白质中有 90% 以上是血红蛋白，因而血红蛋白的 mRNA 十分稳定。成人肝细胞中存在许多编码循环系统所需蛋白质的 mRNA，因而这些 mRNA 也是相对稳定的。

由于终端分化的细胞中 mRNA 极其稳定，再加上 DNA 的扩增和强有力的启动子的转录，真核细胞中有时出现奇迹般的蛋白质合成。家蚕丝心蛋白的合成就是最好的例子。丝心蛋白基因为单拷贝，具有很强的启动子。当调节基因转录的信号出现后，几天功夫便可合成 $10^5$ 丝心蛋白的 mRNA 分子，其平均寿命为几天，比典型真核生物 mRNA 的寿命(约 3 小时)长得多。每个 mRNA 分子可以重复翻译产生 $10^5$ 个丝心蛋白。也就是说，在大约 4 天时间内一个丝心蛋白基因的表达可以产生 $10^{10}$ 个丝心蛋白，二倍体细胞则可产生 $2\times10^{10}$ 个丝心蛋白。家蚕丝腺在发育中经历一系列只有染色体复制而没有细胞分裂的过程，所以成熟丝腺实际上是一个 $10^6$ 倍体的巨大细胞。每套染色体的丝心蛋白基因都活跃地表达。这样，家蚕丝腺在短短四天时间内可以产生 $10^{16}$ 个丝心蛋白分子。mRNA 寿命的延长增加了细胞内某种 mRNA 的浓度，因而提高蛋白质合成起始的速度。因此这种翻译水平的调控形式有时被称作翻译扩增(translational amplification)。在家蚕丝心蛋白合成过程中，存在三个水平上的扩增，即复制水平、转录水平和翻译水平上的三次扩增。

### 二、mRNA 翻译起始的调控

翻译调控的另一个重要途径是控制的 mRNA 可翻译性，也就是控制翻译的起始。许多真核生物的卵细胞都贮藏着暂时不翻译的m RNA。这些不活跃的 mRNA 称为隐蔽 mRNA (masked mRNA)。在受精后几分钟，这些 mRNA 便开始翻译。受精卵中一定存在着激活隐蔽 mRNA 的机制，可能是一种或多种促进 mRNA 与核糖体结合的因子。

非洲爪蟾卵细胞贮藏有大量隐蔽 mRNA，在受精前进行很低水平的蛋白质合成。如果将来自网织红细胞的珠蛋白 mRNA 和放射性标记的氨基酸注射进未受精卵，则有放射性珠蛋白合成。注射进的 mRNA 越多，珠蛋白合成的量也越多，可是同时正常的卵细胞蛋白质合成也下降得越多(图 9-28a)。如果注射含有珠蛋白mRNA的多核糖体，则珠蛋白合成量的增加并不伴随正常卵细胞蛋白质合成量的显著减少(图 9-28b)。

在卵细胞内，核糖体、氨酰基-tRNA，蛋白质合成所需的起始因子，延伸因子和终止因子数量都是足够的，所缺乏的可能就是某些能促进核糖体与 mRNA 结合的因子。这种因子也被称作招募因子(recruitment factor)。

现在人们已经发现，当用各种生长因子刺激哺乳动物细胞时，40S 亚基中的核糖体蛋白质 S6 出现了磷酸化，同时使mRNA上出现了更多的核糖体。受精作用促进了生长因子基因的表

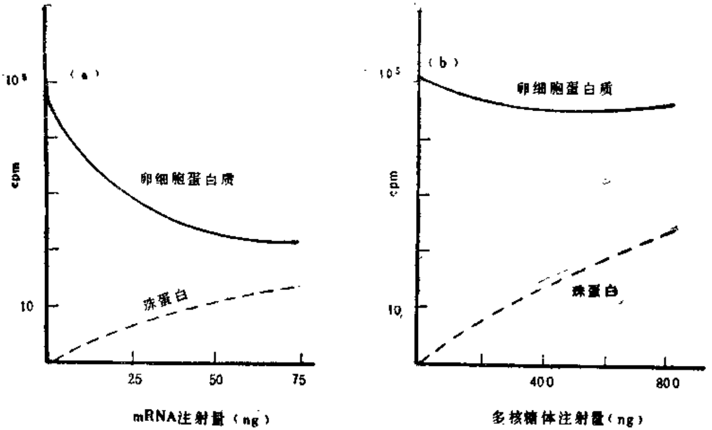



图 9-28 注射网织红细胞珠蛋白 mRNA 对卵细胞蛋白质合成的影响
(a) 注射珠蛋白 mRNA； (b) 注射含有珠蛋白 mRNA 的多核糖体

达，这是很有可能的。核糖体蛋白质的修饰，所改变的是全体核糖体的性能。对于受精后各种隐蔽 mRNA 的翻译的特异性(选择性)可能还需要各别的因子，以促进特定的 mRNA 与核糖体的结合或者是打掉特定 mRNA 上阻止翻译的蛋白质因子。这些可能性都有待于实验的证实。

### 三、真核生物蛋白质合成的自体调控

真核生物的蛋白质合成也像原核生物一样，也有许多自体调控的机制，现在已有许多例证。这里介绍一种原核生物中尚未发现过的自体调控机制，即 β- 微管蛋白和 α- 微管蛋白的异二聚体通过识别新生的 β- 微管蛋白的 N 端几个氨基酸而促进 β-微管蛋白 mRNA 降解的自体调控机制。

真核细胞中的微管是由两种微管蛋白构成的，一种是 α，另一种是 β，它们都是一个不大的基因家族所编码的。除了在发育和分化的不同时刻启动这个基因家族中的不同成员进行表达之外，对于微管蛋白的数量也必然有一个调控机制。人们已经发现细胞中微管蛋白 mRNA 的稳定性与微管蛋白单体的浓度密切相关。当提高细胞内微管蛋白单体浓度时，则与核糖体结合的微管蛋白的 mRNA 稳定性急剧降低。人们通过实验，首先发现这种识别信号存在于前 16 个氨基酸序列中，进一步实验发现只需要 N 端的四个氨基酸即甲硫氨酸-精氨酸-谷氨酸-异亮氨酸。在目前已知氨基酸序列的为数众多的蛋白质中，只有 β-微管蛋白具有这样的四氨基酸末端，也只有微管蛋白(β 和 α)才具有前三个氨基酸的 N 端序列，这可能并非偶然，或许 α-微管蛋白的 mRNA 稳定性受同样的机制所控制。这种机制是 β- 与 α- 微管蛋白结合于刚从核糖体大亚基中冒出的四个氨基酸，然而通过核糖体的"传导"，触发早已与活跃翻译的核糖体相连的一个 RNA 酶，从而切断 mRNA。至于这一 RNA 酶是什么，现在还毫无所知，或许是核糖体结构本身的某种改变。这一调控的模式示于图 9-29。还需指出，这种四氨基酸信号必须位于 N-末端，如果在肽链内部则无效。而且，即使 N-末端是这样的四氨基酸信号，肽链

还必须有一定长度，信号部分才能从核糖体中冒出来。细菌核糖体能掩盖30～40个氨基酸，真核生物核糖体掩盖的肽链可能要长得多。已有的实验证据是，如果阅读框只有26个密码子，则失去对mRNA稳定性的调控。还有，这种形式的自体调控方式只限微管蛋白还是具有某种程度的普遍性，也毫无所知。

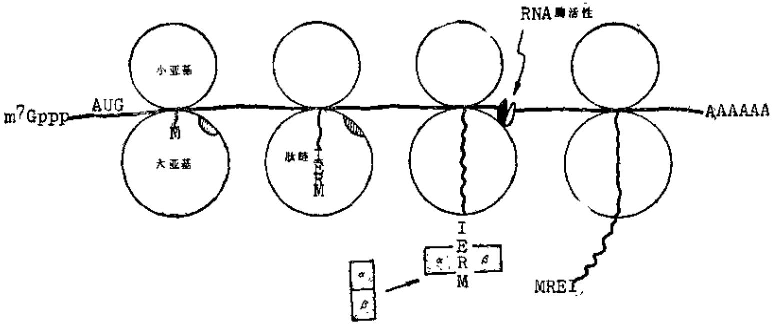



图9-29 β-微管蛋白合成的自体调控是通过控制mRNA的稳定性而实现的

## 提　　要

真核生物的绝大多数都是多细胞的复杂的有机体，其细胞核具有核膜。基因数目多，且大部分含有内元。除了基因序列之外，基因组中还有大量非编码序列。DNA与组蛋白形成核小体结构，同时还与许多非组蛋白的结合。这些结构上的特点决定了调控方面的特点。与原核生物相比，真核生物表达调控的范围更广，花样更多。特别是真核生物有精细的有条不紊的个体发育过程，因而基因表达是严格地按照既定的程序进行的。然而，真核生物和原核生物一样，转录水平的调控是最主要的调控。

真核生物的特定组织中，往往几种或几十种mRNA就要占到mRNA总量的一半以上。这些mRNA称为丰富mRNA。而另外一半的mRNA包括了数千甚至数万种mRNA。这些mRNA称为稀少mRNA。哺乳动物某一细胞类型中只有大约10%的mRNA种类是该细胞类型所特有的；而90%的mRNA种类也同样存在于其它细胞类型。前者叫做奢侈基因，后者叫做持家基因。在个体发育过程中，表达的基因数目逐步减少。在胚胎分化过程中，有可能存在着相对简单的分子开关。果蝇的同型异位现象为此提供了很好的证据。

真核生物在DNA水平上实行调控的例子并不多见。这方面主要有低等真核生物的基因丢失；两栖类和昆虫卵母细胞中rDNA的扩增；基因产物抑制剂所引起的基因扩增；酵母接合型的互变；脊椎动物免疫球蛋白基因的产生等。

活跃转录的基因在染色质结构方面不同于非转录基因。凡是活跃转录的基因均位于常染色质中，处于异染色质中的基因则不表达。现有证据表明，RNA聚合酶Ⅱ不能从有核小体结构的启动子上发起转录，但可以进行链的延伸并同时置换核小体。活跃转录的基因及其两侧都表现出对DNA酶Ⅰ的优先敏感性。如果用极低浓度的DNA酶Ⅰ处理染色质，则可发现一系列DNA酶Ⅰ超敏感位点，其中大部分位于基因5′端的启动子区域，少部分位于其它部位。组蛋

白的修饰作用,如乙酰化,甲基化,磷酸化和泛素化,常常影响染色质结构,从而影响基因的转录活性。DNA 中胞嘧啶的甲基化亦与转录控制有关。

对于真核基因表达的调控,Britten 和 Davidson 于 1969 年提出了类似于原核生物中操纵元或调控元的模型,即 Britten-Davidson 模型,但它仍有待于实验的检验。真核生物 RNA 聚合酶Ⅱ的启动子,除了 TATA 框这一近启动子成分外,还有许多上游启动子成分,如 CCAAT 框,GGGCGG 框等。在启动子的远上游还常有增强子成分。这些顺式调控成分都有各自的蛋白质因子与之结合。这些反式作用因子至少包含两个功能结构域:DNA 结合结构域和转录激活结构域。前者负责与特定的顺式调控成分相结合;而后者则负责与 RNA 聚合酶或通用转录因子相互作用以促进转录。

真核生物RNA 聚合酶Ⅰ只转录主体 rRNA 基因。每种生物都有特定的转录因子与 RNA 聚合酶Ⅰ结合从而促进与启动子形成转录起始复合物。这就决定了 RNA 聚合酶Ⅰ的种族特异性。细胞中 rRNA 的需要量很大。rRNA 基因呈串联重复排列,NTS 序列中有多拷贝启动子序列,从而使 rRNA 基因的转录能够得到串级放大。

真核生物RNA 聚合酶Ⅲ的启动子并非都是基因内启动子,U6 和 7SK 的启动子完全在转录起始点的上游,而且有 TATA 样序列。真核生物的三种 RNA 聚合酶可能使用共通的转录因子,而且还有共通的亚基。这些都暗示着,三种转录机构之间可能存在着某种协调机制。

真核生物由于核膜的存在,是不可能存在着衰减子调控的。然而,翻译对于转录的间接控制机制仍然存在。例如酵母的氨基酸通用控制机制中的反式作用因子GCN4 mRNA的前导序列中,有四个很短的开放阅读框架,从翻译水平上控制 GCN4 基因的表达,再通过 GCN4 控制氨基酸合成基因的转录。

真核生物转录后水平的调控的余地要比原核生物大得多,这一方面是由于核膜的存在,必须有一个 hnRNA 选择性加工运输的问题,另一方面由于真核基因内元的存在,必然存在着复杂的拼接过程并产生某些场合下的选择性拼接。

真核生物在翻译水平上的调控也是值得注意的。真核细胞有许多因素能增加需要量大的蛋白质 mRNA 的寿命,如乳腺中的酪蛋白 mRNA,组织红细胞中的血红蛋白 mRNA 等。与此相反,一些决定细胞周期的蛋白质 mRNA 的寿命则很短。在真核生物中还存在着 mRNA 可翻译性的控制机制。卵细胞中存在着许多暂时不翻译的隐蔽 mRNA,受精后便开始翻译。可能受精过程使某些因子得到表达,这些因子促进了核糖体与 mRNA 的结合。真核生物的蛋白质的合成也存在着许多自体调控的机制,其中微管蛋白合成的自体调控是颇为独特的。

## 参 考 文 献


1. Britten, R. J. and Davidson, E. H. (1969) "Gene Regulation for Higher Cells: A Theory" *Science* **165**, 349～357.
2. Lewin, B. (1980) *Gene Expression 2, Eucaryotic Genomes,* Wiley, New York.
3. Davidson, E. H. (1986) *Gene Activity in Early Development,* 3rd ed., New York, Acad. Press.
4. Scott, M. P. (1985) "Molecules and Puzzles from the Antennapedia Homeotic Gene Complex of *Drosophila.*" *Trends in Genetics* 1: 74～80.
5. Ingham, P. (1985) "The Regulation of the Bithorax Complex," *Trends in Genetics* 1: 112～116.
6. Wright, C. V. E. et al. (1989) "Vertebrate Homeodomain Proteins, Families of Region-Specific Transcription Factors." *Trends. Biochem. Sci.* 14: 52～56.
7. Stark, G. R. and Wahl, G. M. (1984) "Gene Amplification." *Ann. Rev. Biochem.* 53: 447～491.

8. Hourcade, D. et al. (1973) "The Amplification of Ribosomal RNA Genes Involves a Rolling Circle Intermediate." *Proc. Nat. Acad. Sci.* 70: 2926～2930.
9. De Cicco, D. V. and Spradling, A. C. (1984) "Localization of a Cis-acting Element Responsible for the Developmentally Regulated Amplification of *Drosophila* Chorion genes." *Cell* 38: 45～54.
10. Yao, M.-C. et al. (1985) "Gene Amplification in *Tetrahymena thermophila*: Formation of Extrachromosomal Palindromic Genes Coding for rRNA." *Mol. Cell. Biol.* 5: 1260～1267.
11. Schimke, R. (1984) "Gene Amplification in Cultured Animal Cells." *Cell* 37: 705～713.
12. Strathern, J. N. et al. (1980) "The Structure and Organization of the Transposable Mating Type Cassettes in *Saccharomyces*." *Proc. Nat. Acad. Sci.* 77: 2839～2843.
13. Astell, C. et al. (1981) "The Sequence of the DNAs Coding for Mating-type Loci of *Saccharomyces cerevisiae*." *Cell* 27: 15～23.
14. Nasmyth, K. (1982) "Regulation of Yeast Mating-type Chromatin Structure by SIR: An Action at a Distance Affecting Both Transcription and Transposition." *Cell* 30: 567～578.
15. Klar, A. J. S. (1987) "Determination of the Yeast Cell Lineage." *Cell* 49: 433～435.
16. Bender, A. and Sprague, G. F. (1987) "MATα1 Protein, a Yeast Transcription Activator, Binds Synergistically with a Second Protein to a Set of Cell-type Specific Genes." *Cell* 50: 681～691.
17. Nasmyth, K. et al. (1987) "Both Positive and Negative Regulatorsof HO Transcription are Required for Mother Cell Specific Mating Type Switching in Yeast." *Cell* 48: 579～587.
18. Maizels, N. (1987) "Diversity Achieved by Diverse Mechanisms: Gene Conversion in Developing B Cell of the Chicken." *Cell* 48: 359～360.
19. Golub, E. S. (1987) "Somatic Mutation: Diversity and Regulation of the Immune Repertoire." *Cell* 48: 723～724.
20. Felsenfeld, G. and McGhee, J. D. (1986) "Structure of the 30 nm Chromatin Fiber." *Cell* 44: 375～377.
21. Knezetic, J. A. and Luse, D. S. (1986) "The Presence of Nucleosomes on a DNA Template Prevents Initiation by RNA Polymerase II in Vitro." *Cell* 45: 95～104.
22. Ramain, P. et al. (1986) "Changes in the Chromatin Structure of *Drosophila* Glue Genes Accompany Developmental Cessation of Transcription in Wild Type and Transformed Strains." *Cell* 45: 545～553.
23. Weisbrod, S. (1982) "Active Chromatin." *Nature* 297: 289～294.
24. Elgin, S. C. R. (1981) "DNAase I-hypersensitive Sites of Chromatin." *Cell* 27: 413～415.
25. Wu, C. (1980) "The 5′ends of *Drosophila* Heat Shock Genes in Chromatin are Hypersensitive to DNAase I." *Nature* 286: 854～860.
26. Saragosti, S. et al. (1980) "Absence of Nucleosomes in a Fraction of SV40 Chromatin between the Origin of Replication and the Region Coding for the Late Leader RNA." *Cell* 20: 65～73.
27. Elgin, S. C. R. (1984) "Anatomy of Hypersensitive Sites." *Nature* 309: 213～214.
28. Varshavsky, A. et al. (1982) "Cellular and SV40 Chromatin: Replication, Segregation, Ubiquitination, Nuclease-hypersensitive Sites, HMG-Containing Nucleosomes, and Heterochromatin-Specific Protein." *Cold Spring Harbor Symp. Quant. Biol.* 47: 511～520.
29. Lorch, Y. et al. (1987) "Nucleosomes Inhibit the Initiation of Transcription but Allow Chain Elongation with the Displacement of Histones." *Cell* 49: 203～210.
30. Allfrey, V. G. (1977) "Post-synthetic Modification of Histone Structure: A Mechanism for the Control of Chromosome Structure by the Modulation of Histone-DNA Interactions." in *Chromatin And Chromosome Structure*, ed. Li, H. J. and Eckhardt, R., Academic Press, New York, pp. 167～191.
31. Keene, M. A. et al. (1984) "Some Observations on DNA Structure and Chromatin Organization at Specific Loci in *Drosophila melanogaster*." in *Eukaryotic Gene Expression*, ed. Kumar, A., Plenum Press, New York, pp. 1～25.
32. Cedar, H. (1984) "DNA Methylation and Gene Expression." in *DNA Methylation*, ed. Razin, A. et al., Springer-Verlag, New York, pp. 147～164.
33. Doerfler, W. (1983) "DNA Methylation and Gene Activity." *Ann. Rev. Biochem* 52: 93～124.
34. Yaniv, M. et al. (1986) "Structure of Transcriptionally Active Chromatin." *CRC Crit. Rev. Biochem.* 21: 1～22.
35. Gross, D. S. and Garrard, W. T. (1988) "Nuclease Hypersensitive Sites in Chromatin." *Ann. Rev. Biochem.*

57: 159～197.
36. Mitchell, P. J. and Tjian, R. (1989) "Transcriptional Regulation in Mammalian Cells by Sequence-Specific DNA Binding Proteins." Science **245**: 371～378.
37. McKnight, S. L. and Kingsbury, R. (1982) "Transcriptional Control Signals of a Eukaryotic Protein-Coding Gene." *Science* 217: 316～324.
38. Cochran, M. D. and Weissmann, C. (1984) "Modular Structure of the β-Globin and the TK Promoters." *EMBO J.* 3: 2453～2459.
39. McKnight, S. L. et al. (1984) "The Distal Transcription Signals of the Herpes Virus tk Gene Share a Common Hexanucleotide Control Sequence." *Cell* 37: 253～262.
40. Khoury, G. and Gruss, P. (1983) "Enhancer Elements." *Cell* 33: 313～314.
41. Scholer, H. R. and Gruss, P. (1984) "Specific Interaction Between Enhancer-Containing Molecules and Cellular Components." *Cell* 36: 403～411.
42. Schaffner, W. et al. (1985) "Enhancers and Eukaryotic Gene Transcription." *Trends in Genetics* 1: 224～230.
43. Sassone-Corsi, P. et al. (1985) "A *Trans*-Acting Factor is Responsible for the Simian Virus 40 Enhancer Activity *in Vitro*. *Nature* 312: 458～463.
44. Robertson, M. (1987) "Gene Regulation: Specificity and Flexibility." *Nature* 327: 464～466.
45. Dynan, W. S. and Tjian, R. (1985) "Control of Eukaryotic Messenger RNA Synthesis by Sequence-Specific DNA Binding Proteins." *Nature* 316: 774～778.
46. Gidoni, D. et al. (1984) "Multiple Specific Contacts Between a Mammalian Transcription Factor and Its Cognate Promoters." *Nature* 312: 409～413.
47. Fire, A. et al. (1984) "Interactions Between RNA Polymerase II, Factors, and Template Leading to Accurate Transcription." *J. Biol. Chem.* 259: 2509～2516.
48. Maniatis, T. et al. (1987) "Regulation of Inducible and Tissue-Specific Gene Expression." *Science* 236: 1237～1245.
49. Ondek, B. et al. (1988) "The SV40 Enhancer Contains Two Distinct Levels of Organization." *Nature* 333: 40～45.
50. Sen, R. and Baltimore, D. (1986) "Multiple Nuclear Factors Interact with the Immunoglobulin Enhancer Sequences." *Cell* 46: 705～716.
51. Kadonaga, J. T. et al. (1987) "Isolation of cDNA Encoding Transcription Factor Sp1 and Functional Analysis of the DNA Binding Domain." *Cell* 51: 1079～1090.
52. Jones, K. A. et al. (1987) "A Cellular DNA-Binding Protein That Activates Eukaryotic Transcription and DNA Replication." *Cell* 48: 79～89.
53. Dorn, A. et al. (1987) "A Multiplicity of CCAAT Box Binding Proteins." Cell 50: 863～872.
54. Fletcher, C. et al. (1987) "Purification and Characterization of OTF-1, a Transcription Factor Regulating Cell Cycle Expression of a Human Histone H2b Gene." *Cell* 51: 773～781
55. Scheidereit, C. et al. (1987) "Identification and Purification of a Human Lymphoid-Specific Octamer Binding Protein (OTF-2) That Activates Transcription of an Immunoglobulin Promoter In Vitro." *Cell* 51: 783～793.
56. Carbon, P. et al. (1987) "A Common Octamer Motif Binding Protein Is Involved in the Transcription of U6 snRNA by RNAPolymerase III and U2 snRNA by RNA Polymerase II." *Cell* 51: 71～79.
57. Imagawa, M. (1987) "Transcription Factor AP-2 Mediates Induction by Two Different Signal Transduction Pathways: Protein Kinase C and cAMP." *Cell* 51: 241～249.
58. Zheng, X.-M. et al. (1987) "A General Transcription Factor Forms a Stable Complex with RNA Polymerase B (II)." *Cell* 50: 361～368.
59. Ma, J. and Ptashne, M. (1987) "A New Class of Yeast Transcriptional Activators." *Cell* 51: 113～119.
60. Bohmann, D. et al. (1987) "Human Proto-oncogene *c-jun* Encodes a Protein with Antigenic and Enhancer Binding Properties of Transcription Factor AP-1." *Science* 238: 1386～1392.
61. Struhl, K. (1988) "The JUN Oncoprotein, a Vertebrate Transcription Factor, Activates Transcription in Yeast." *Nature* 332: 649～650.
62. Lee, W. et al. (1987) "Purified Transcription Factor AP1 Interacts with TPA-Inducible Enhancer Elements."

*Cell* 49: 741～752.

63. Maki, Y. et al. (1987) "Avian Sarcoma Virus 17 Carries the *jun* Oncogene." *Proc. Nat. Acad. Sci.* 84: 2848～2852.

64. Struhl, K. (1987) "The DNA-Binding Domain of the JUN Oncoprotein and Yeast GCN4 Transcriptional Activator Protein Are Functionally Homologous." *Cell* 50: 841～846.

65. Bos, T. J. et al. (1988) "*v-jun* Encodes a Nuclear Protein with Enhancer Binding Properties of AP-1." *Cell* 52: 705～712.

66. Sommerville, J. (1984) "RNA Polymerase I Promoters and Transcription Factors." *Nature* 310: 189～190.

67. Miesfeld, R. and Arnheim, N. (1984) "Species-Specific rDNA Transcription Is Due to Promoter-Specific Binding Factors." *Mol. Cell. Biol.* 4: 221～227.

68. Learned, R. M. et al. (1985) "Purification and Characterization of a Transcription Factor That Confers Promoter Specificity to Human RNA Polymerase I." *Mol. Cell. Biol.* 5: 1358～1369.

69. Bateman, E. and Paule, M. R. (1986) "Regulation of Eukaryotic Ribosomal RNA Transcription by RNA Polymerase Modification." *Cell* 47: 445～450.

70. Kownin, P. et al. (1987) "Eukaryotic RNA Polymerase I Promoter Binding Is Directed by Protein Contacts with Transcription Initiation Factor and Is DNA Sequence-Independent." *Cell* 50: 693～699.

71. Baker, S. M. and Platt, T. (1986) "Pol I Transcription: which Comes First, the End or the Beginning?" *Cell* 47: 839～840.

72. Henderson, S. and Sollner-Webb, B. (1986) "A Transcriptional Terminator Is a Novel Element of the Promoter of the Mouse Ribosomal RNA Gene." *Cell* 47: 891～900.

73. Grummt, I. et al. (1986) "A Transcription Terminator Located Upstream of the Mouse rDNA Initiation Site Affects rRNA Synthesis." *Cell* 47: 901～911.

74. McStay, B. and Reeder, R. H. (1986) "A Termination Site for *Xenopus* RNA Polymerase I Also Acts as an Element of an Adjacent Promoter." *Cell* 47: 913～920.

75. Sakonju, S. et al (1980) "A Control Region in the Center of the 5S RNA Gene Directs Specific Initiation of Transcription I: The 5' Border of the Rigion." *Cell* 19: 13～25.

76. Bogenhagen, D. F. et al. (1980) "A Control Region in the Center of the 5S RNA Gene Directs Specific Initiation of Transcription II: The 3' Border of the Region." *Cell* 19: 27～35.

77. Murphy, S. et al. (1987) "The In Vitro Transcription of the 7SK RNA Gene by RNA Polymerase III Is Dependent Only on the Presence of an Upstream Promoter." *Cell* 51: 81～87.

78. Sollner-Webb, B. (1988) "Surprises in Polymerase III Transcription." *Cell* 52: 153～154.

79. Geiduschek, E. P. (1988) "Transcription by RNA Polymerase III." *Ann. Rev. Biochem.* 57: 873～914.

80. Ilan, J. ed. (1987) *Translational Regulation of Gene Expression*, Plenum Press, New York.

81. Mueller, P. P. and Hinnebusch, A. G. (1986) "Multiple Upstream AUG Codons Mediate Transcriptional Control of GCN4." *Cell* 45: 201～207.

82. Hinnebusch, A. G. and Mueller, P. P. (1987) "Translational Control of a Transcriptional Activator in the Regulation of Amino Acid Biosynthesis in Yeast." in *Translational Regulation of Gene Expression*, ed. Ilan, J., Plenum Press, New York, pp. 397～412.

83. Fink, G. R. (1986) "Translational Control of Transcription in Eukaryotes." *Cell* 45: 155～156.

84. Darnell, J. E., Jr. (1983) "The Processing of RNA." *Sci. Amer.* 249: 89～99.

85. Beyer, A. L. et al. (1977) "Identification and Characterization of Packaging Proteins of Core 40S hnRNP Particles." *Cell* 11: 127～138.

86. Nevins, R. J. (1983) "The Pathway of Eukaryotic mRNA Formation." *Ann. Rev. Biochem.* 52: 441～466.

87. Birnstiel, M. L. et al. (1985) "Transcription Termination and 3' Processing: The End Is in Site!" *Cell* 41: 349～359.

88. Proudfoot, N. J. (1989) "How RNA Polymerase II Terminates Transcription in Higher Eukaryotes." *Trends Biochem. Sci.* 14: 105～110.

89. Breitbart, R. E. (1987) "Alternative Splicing: A Ubiquitous Mechanism for the Generation of Multiple Protein Isoforms from Single Genes." *Ann. Rev. Biochem.* 56: 467～495.

90. Laz, T. et al. (1987) "The Role of Messenger RNA Sequence and Structure in Eukaryotic Translation." in *Translational Regulation of Gene Expression*, ed. Ilan, J., Plenum Press, New York, pp. 413～429.

91. Brawerman, G. (1987) "Determinants of Messenger RNA Stability." *Cell* 48: 5~6.
92. Candelas, G. C. et al. (1987) "Strategies of Fibroin Production." in *Translational Regulation of Gene Expression*, ed. Ilan, J., Plenum Press, New York, pp. 209~228.
93. Richter, J. D. (1987) "Molecular Mechanisms of Translational Control during the Early Development of *Xenopus laevis*." in *Translational Regulation of Gene Expression*, ed. Ilan, J., Plenum Press, New York, pp. 111~139.
94. Yen, T. J. et al. (1988) "Autoregulated Instability of β-Tubulin mRNAs by Recognition of the Nascent Amino Terminus of β-Tubulin." *Nature* 334: 580~585.
95. Cleveland, D. W. (1988) "Autoregulated Instability of Tubulin mRNAs." *Trends Biochem. Sci.* 13: 339~343.

# 第十章　基因表达调控的分子机制

## 第一节　概　　述

基因按其最终产物可以分为 RNA 基因和蛋白质基因两种类型。对于 RNA 基因的表达，它只需要转录一个过程；而对于蛋白质基因的表达，则需要转录和翻译两个过程。从广义来说，细胞内任何一个生理生化反应都会给基因表达带来影响。把这些影响都划归基因表达的调控范畴，显然是没有意义的。因此基因表达调控的严格定义应该是：任何影响转录过程和翻译过程的开启关闭和这两个过程的速率的较为直接的因素及其作用，就叫做基因表达的调控。在这一定义中，突出的要点就是四个字：开、关、快(多)，慢(少)。一般来说，蛋白质基因产生原初的翻译产物之后，常需要进行翻译后处理和修饰。这一过程是基因产物的定位和发挥其生理功能所不可缺少的。这一过程所需的信息也是基因信息的转化形式，即由核苷酸序列转变为氨基酸序列。然而在一般情况下这一过程具有很高程度的非逆转性，因而在基因调控方面的意义也就小得多。这样，遗传信息流动过程的绝大部分，从 DNA 合成，RNA 合成直到蛋白质合成，都是基因表达调控的舞台。其中转录过程的调控最为重要，原核生物的基因表达的调控是这样，真核生物的基因表达的调控也是这样。原核生物和真核生物在基因表达的调控方面尽管各有各的花样，但二者具有惊人的平行关系和相似性。说起来这也并不奇怪。一方面，原核生物和真核生物具有共同的祖先；另一方面，人们相信，在分子水平的作用机理上，二者是相近的。

在基因表达调控的舞台上，主要有三类表演动作：一是核苷酸序列之间的相互作用，包括核酸分子内部核苷酸序列之间的相互作用和两个核酸分子之间核苷酸序列的相互作用；二是核酸分子的核苷酸序列与蛋白质(酶)分子之间的相互作用；三是蛋白质(酶)分子之间的相互作用。在这三种相互作用中，第二种作用居于首要地位，是基因表达调控的核心问题。其余两种作用也常常借助于第二种作用才能实现对基因表达的调控。第一种作用主要是通过互补碱基之间的氢键作用而实现的；第三种作用虽很重要并已有所了解，但可能花样繁多。本章着重于阐述第二种相互作用的性质、特征及其细微机制。然后，我们将讨论结合于 DNA 链的不同位点上的蛋白质相互作用中可能存在的机制。最后再简略地阐述 RNA 所组成的开关系统。

## 第二节　两类遗传信息的差异[1~2]

前面我们已经讲到过 C 值矛盾，即人们根据真核生物的基因平均大小，可能的基因数目以及其他有关现象(如介入序列等)只能够解释基因组大小的 10%，考虑某些难以觉察的非必需基因，基因所占基因组 DNA 含量的百分比最多也不会超过 20%。其余 80% 的 DNA 到底起什么作用？现在还不可能做出令人满意的解释。但是可以肯定的一点是，基因组中大部分 DNA 序列是用来编码基因选择性表达的遗传信息的。也就是说这部分 DNA 贮藏了基因选择性表达的指令。这类遗传信息我们称之为第Ⅱ类遗传信息，而经典的三联体密码子我们称

之为第Ⅰ类遗传信息。

这两类遗传信息有着很大的不同。首先，第Ⅰ类遗传信息强烈地依赖于外在的蛋白质，酶和小分子核酸所组成的复杂体系，如转录体系，蛋白质合成机构等；而其序列本身的差异只是提供了外在系统识别的基础。因而，第Ⅰ类遗传信息的解读机制是外在的。第Ⅱ类遗传信息除了需要外在的系统与之作用之外，还会“主动”修饰自己的双螺旋空间结构，以便实现对基因表达的调控。因而，第Ⅱ类遗传信息的解读是以内在和外在相结合的方式进行的。其次，第Ⅰ类遗传信息在解读时涉及的范围较小，而第Ⅱ类遗传信息在解读时能够引起较大范围的双螺旋结构的变化，这种变化可能是以低频振动的方式进行传播的。第三，第Ⅰ类遗传信息在解读时需要按照 Watson 规律形成碱基配对而同时中断原有的双螺旋之间的碱基配对；而第Ⅱ类遗

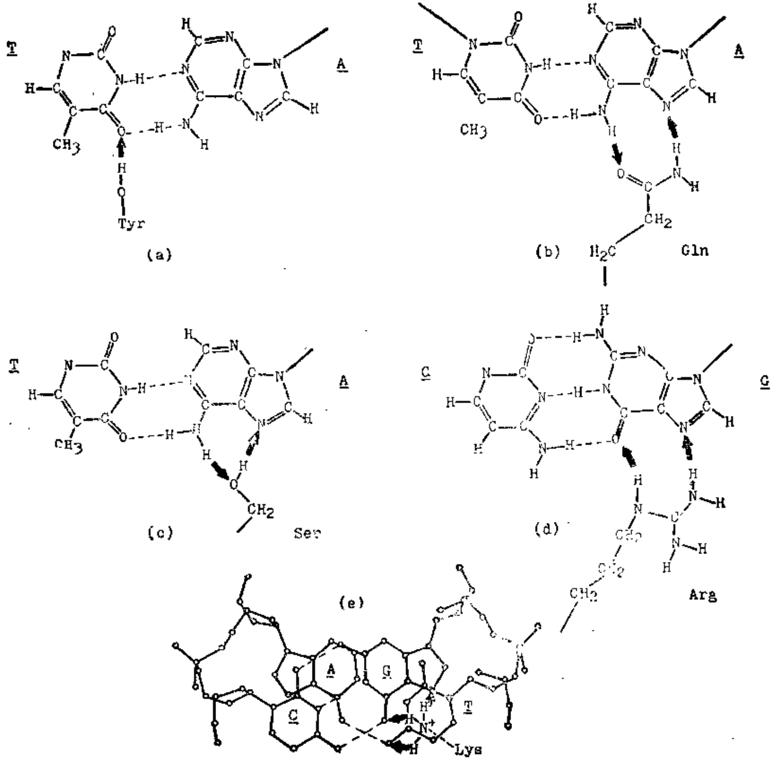



图 10-1 (a) Tyr OH 基与 T 的 $O_4$ 形成氢键
(b) Gln 的酰基和氨基与 A 的 $N_6$ 和 $N_7$ 形成两个氢键
(c) Ser 的 OH 基与 A 的 $N_6$ 和 $N_7$ 形成两个氢键
(d) Arg 的胍基与 G 的 $O_6$ 和 $N_7$ 形成两个氢键
(e) Lys 的 ε- 氨基分别与 T 的 $O_4$ 和 G 的 $O_6$ 形成两个氢键，还有可能与 G 的 $N_7$ 形成第三个氢键。

传信息的解读并不需要中断原有的双螺旋之间的碱基配对，而是通过调控蛋白质的氨基酸侧链与大沟(或小沟)中的碱基对边缘的潜在的氢原子供体或受体之间形成氢键（见第一章第七节)。第四，第Ⅰ类遗传信息的编码方式是一维的、线性的，只位于一条 DNA 链上；而第Ⅱ类遗传信息的编码方式是三维的、空间的，可以位于两条 DNA 链上。因此第Ⅱ类遗传信息的密码又称为空间密码或调控密码。第五，调控密码可能具有更高的简并性。例如一种碱基对排列花式可能与不同的氨基酸侧链发生相互作用；一种氨基酸侧链也可能与不同的碱基对排列花式发生相互作用。

由于上述种种原因，尽管多年以前就有人提出调控密码的概念，但这种密码的具体形式仍然不清楚，有时一种氨基酸侧链与一个碱基发生作用，例如酪氨酸的 OH 基能与胸腺嘧啶的 $O_4$ 相互作用生成一个氢键(图 10-1a)。有时一个氨基酸侧链能与一个碱基相互作用而生成两个氢键，例如谷酰胺与腺嘌呤之间、丝氨酸与腺嘌呤之间、精氨酸与鸟嘌呤之间(图 (10-1 b, c, d)。有时一个氨基酸侧链能同时与两个碱基发生相互作用，例如赖氨酸的 ε- 氨基能与相邻的鸟嘌呤和胸腺嘧啶相互作用生成三个氢键(图 10-1e)。氨基酸侧链与碱基之间不但可以形成氢键，而且可以通过 van der Waal 力相互作用。因此二者作用方式极为复杂，实际上并不存在几联体密码子的问题。

然而，人们通过研究具体的调控蛋白质与其作用位点(DNA)的相互关系，确实发现了许多很有价值的规律，充实和发展了调控密码的概念。

## 第三节　调节蛋白质的对称性与其识别序列的对称性[3~4]

调节蛋白质与酶分子一样，都是由许多较短的 α 螺旋，β 折叠再由许多转角连接起来的成为球状或近似于球状的蛋白质分子。很长的 α 螺旋或 β 折叠所产生的刚性结构是很难与 DNA 分子发生相互作用的。大部分的调控蛋白质都是同种分子的多聚体，尤其是以二聚体为多，其次是四聚体。调控蛋白质以多聚体形式与 DNA 相互作用，除了具有通常所说的节省编码的 DNA 和减少大分子装配的随机错误等优点以外，更重要的优点是通过单体和多聚体之间的平衡而达到迅速开启或关闭基因活性的目的。对于二聚体来说，其浓度是随着单体浓度的平方而变化的，因此由于单体的合成或破坏而迅速地影响到 DNA 分子上二聚体的结合与解离。还有一个重要的优点是，调控蛋白质以同种二聚体的形式出现，则其识别位点也必须以某种重复序列与之适应。这样，就极大地提高了调控作用的特异性，这对于精细的生命活动是必不可少的。

由于调控蛋白质采取多聚体的形式，就必然存在着分子对称性的问题。由于蛋白质分子是一个相对独立的结构，因而多聚体的可能组合方式较多；而 DNA 的重复序列是在同一条双螺旋分子中，因而 DNA 重复序列的种类是极为有限的。在图 10-2 中列出了四种蛋白质二聚体的组合方式：(1)平移对称；(2)二倍轴对称；(3)二倍旋转对称；(4)螺丝对称。图的右边写出了似乎可能的 DNA 序列安排。实际上只有(1)和(3)两组，蛋白质二聚体的对称性与 DNA 序列的对称性相同。实际上在同一 DNA 分子上两个靠近的 DNA 序列之间是不可能出现第(2)和第(4)种蛋白质二聚体那样的对称性。虽然在第(1)组和第(3)组中，蛋白质二聚体的对称性与 DNA 序列的对称性相符，但是目前已研究过的调控蛋白质与其识别点之间的关系绝大多数属于第(3)组类型。而实际上调控蛋白质的二聚体也都采取二倍旋转对称的

形式，即使是四聚体，它具有二倍旋转对称的特征(图 10-3)。因此，调控蛋白质识别的 DNA 序列必须是反向重复序列。

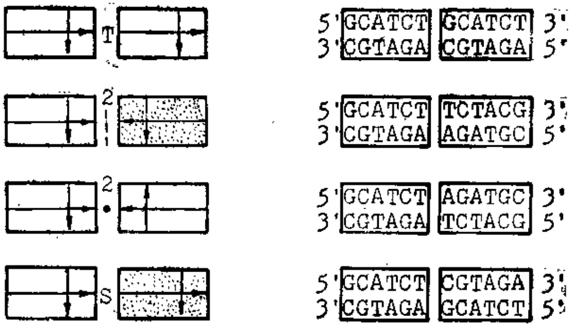



图 10-2　蛋白质二聚体的对称性与可能的 DNA 序列之间的关系

图中空白框表示蛋白质正面，带点框表示其背面，不过这样的规定是任意的。T 表示平移对称，S 表示螺丝对称，2 表示两倍轴对称，2 表示二倍旋转对称。

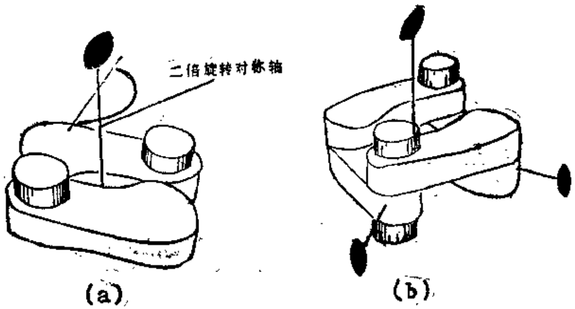



图 10-3　调控蛋白质二聚体和四聚体的右鞋模型

## 第四节　λ 噬菌体的阻遏蛋白和 Cro 蛋白所组成的转录开关系统[5~8]

当 λ 噬菌体处于原噬菌体状态时，由于 *cI* 基因的产物阻遏蛋白结合于左右两个早期操纵元的操作子上，关闭了所有的早期基因和晚期基因的活性，而同时促成了 *cI* 基因的自控表达。*cI* 基因的表达，不但维持了溶原状态，而且产生的 λ 阻遏蛋白作为一种免疫性物质，保护溶原菌免受其它 λ 噬菌体的重复感染。而当溶原菌受到紫外线等作用引起 DNA 损伤时，细胞中的 recA 蛋白便获得蛋白酶的活性，专一性地水解若干调控蛋白质的 Ala-Gly 之间肽键，其中包括 λ 阻遏蛋白。这样，由于 λ 阻遏蛋白的单体和二聚体之间的平衡，λ 阻遏蛋白迅速地从操作子上解离下来。这时 RNA 聚合酶才能进入两个早期操纵元的启动子，首先转录并翻译

出抗终止蛋白 N 和另一个调节蛋白质 Cro。由于 N 的抗终止作用，迟早期基因才得以表达，产生 λ 噬菌体复制所必需的蛋白质 O 和 P 以及另一个抗终止蛋白 Q。由于抗终止蛋白Q的作用晚期基因才得以表达，产生噬菌体的结构蛋白质及裂解细菌和组装噬菌体颗粒所必需的酶。这就是裂解生长周期。再说裂解生长过程中首先表达的另外一个蛋白质 Cro 的作用。Cro 蛋白与 λ 阻遏蛋白不同，λ 阻遏蛋白不但有阻遏转录的作用(负调控)，而且有激活其自身基因(*cI*) 转录的作用(正调控)；而 Cro 蛋白只有负调控作用，但是它首先阻遏 *cI* 基因的转录，然后才阻遏早期基因包括 Cro 自身的基因的转录。

λ 阻遏蛋白和 Cro 蛋白是怎样控制基因活性的开启和关闭的呢？它们在两个早期操纵元的操作子各有三个结合位点，每个结合位点均为 17bp 的DNA 序列(图10-4)，这些序列相似而不相同。正是这些序列的差异，使得每个位点与 λ 阻遏蛋白和 Cro 蛋白的亲和力不同。

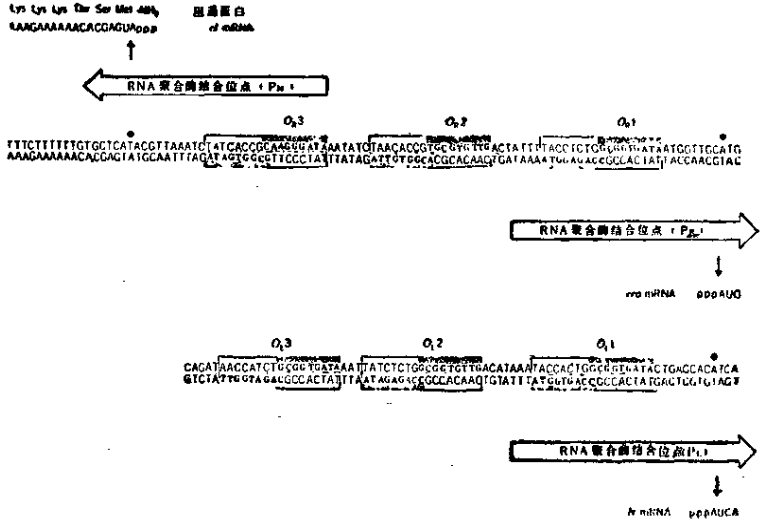



图 10-4 每个早期操纵元的操作子均有三个阻遏物结合位点，其中 $O_{L1}$ 与 $P_L$ 重叠，$O_{R1}$ 与 $P_R$ 重叠，$O_{R3}$ 与 $P_M$ 重叠，为了与 $O_R$ 比较，$O_LP_L$ 的序列及转录方向均已颠倒过来

λ 阻遏蛋白是由 236 个氨基酸组成的多肽链，分子量为 27KDa。λ 阻遏蛋白分子有两个结构域，一个是 N 末端结构域，由第 1 至 92 氨基酸残基所构成，它提供与操作子相结合的位点。另一个是 C 末端结构域，由第 132～236 氨基酸残基所构成，其功能是负责二聚体的形成。两个结构域由第 93～131 氨基酸残基所组成的肽链相连接。这个连结部位很容易被蛋白水解酶如枯草杆菌蛋白酶和木瓜蛋白酶等所切断，使两个结构域分离。分离的 N 末端结构域仍然能够与 DNA 上的 O位点相结合，C 末端结构域相互之间仍然能够形成二聚体，只不过两种作用都没有完整的二聚体的作用强。由此可知，λ 阻遏蛋白分子依赖其 C 末端结构域形成二聚体，又依赖其 N 末端结构域与 O 位点相结合(图 10-5)。这种二聚体与 O 位点结合力大小顺序是 $O_{R1}>O_{R2}>O_{R3}$，$O_{L1}>O_{L2}>O_{L3}$。

Cro 蛋白分子很小，由 66 个氨基酸残基所组成，整个分子只有一个结构域，它既有与DNA的结合位点，又有相互结合成二聚体的结合位点(图 10-5)。该二聚体与 O 位点的亲和力顺序恰恰相反，是 $O_{R3}>O_{R2}>O_{R1}$，$O_{L3}>O_{L2}>O_{L1}$。

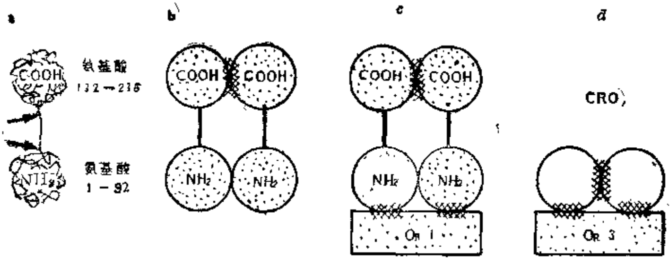



图 10-5　λ 阻遏蛋白和 Cro 蛋白的分子特征

(a) λ 阻遏蛋白的连接部位很容易被蛋白水解酶切断产生两个分离的结构域。

(b) λ 阻遏蛋白依靠 C 末端结构域相互结合成二聚体。为醒目起见，将二聚体画成哑铃状。

(c) λ 阻遏蛋白二聚体依赖其 N 末端结构域与 DNA 结合，与 $O_{R1}$ 和 $O_{L1}$ 结合能力最强。

(d) Cro 蛋白二聚体与 $O_{R3}$ 的结合能力最强

λ 阻遏蛋白与 Cro 蛋白在与 O 位点结合方面，除了上述的结合力顺序不同以外，还有许多重大差别。为了方便起见，我们只讨论他们与 $O_R$ 的结合。首先，λ 阻遏蛋白二聚体与 O 位点的结合具有协同效应。当一个二聚体与 $O_{R1}$ 结合以后，就很容易发生第二个二聚体与 $O_{R2}$ 的结合。这种协同效应也是通过二聚体的 C 末端结构域而实现的。因为 $O_{R1}$ 上的二聚体能促进 $O_{R2}$ 上的二聚体的结合，而 $O_{R2}$ 上的二聚体却不能促进 $O_{R3}$ 上二聚体的结合。通常 $O_{R3}$ 的结合是独立进行的。但是，当 $O_{R1}$ 位点发生突变时，则 $O_{R2}$ 的结合促进了 $O_{R3}$ 的结合，这也说明协同效应是由二聚体 C 末端结构域产生的(图 10-6)。在溶原状态下 $O_{R3}$ 一般是空着的，只有 $O_{R1}$ 和 $O_{R2}$ 结合了两个 λ 阻遏蛋白二聚体，使得 RNA 聚合酶不能进入 $P_R$，遏制了早期基因的转录。其次，由于结合在 $O_{R2}$ 上的二聚体的 *N* 末端结构域与 RNA 聚合酶发生相互作用，或者是促进 RNA 聚合酶与 $P_M$ 结合，或者是促进开放性启动子复合物的形成，从而促进了 *cI* 基因的转录，这就是 λ 阻遏蛋白的正调控作用。影响正调控作用的突变点都位于 α 螺旋 2 以及 α-螺旋 2 与 α 螺旋 3 之间的转角部分。(关于阻遏蛋白的 α 螺旋结构在下一节中讲述)。在正常的阻遏蛋白分子上，这些位点与 RNA 聚合酶接触，突变了的氨基酸增加了分子表面的正电荷，从而不利于这种相互作用。λ 阻遏蛋白和 RNA 聚合酶还同时与一个磷酸基团(共享磷酸)结合。根据共享磷酸的位置不同，Ptashne 等人预言了 $P_{22}$ 阻遏蛋白与 RNA 聚合酶的作用位点在 α 螺旋 3 上。后来对于不能激活 RNA 聚合酶的突变 $P_{22}$ 阻遏蛋白进行序列分析，突变点确实位于 α 螺旋 3 上(图 10-7)。1988 年，Ptashne 等人又按照 λ 阻遏蛋白对于 $P_M$ 正调控所必须的氨基酸(α-螺旋 2 及转角上)重新设计 Cro 蛋白的 α-螺旋 2 及转角(Thr17→Glu，Lys21→Asp，Asp22→Lys，Tyr26→Gly)。这种改造过的 Cro 蛋白称为 Cro67。Cro67 保持了天然 Cro 蛋白所具有的操作子结合能力。若将 $O_{R3}$ 的部分碱基突变，使之丧失与调控蛋白质结合的能力。在这种情况下，Cro67 只能结合于 $O_{R2}$ 和 $O_{R1}$ 并

具有刺激 $P_M$ 进行转录的能力。这就完全证明了 λ 阻遏蛋白 α- 螺旋 2 上的两个酸性氨基酸残基和转角上的 Gly 是直接与 RNA 聚合酶相互作用的(Lys是与DNA作用所必需)。

然而，λ 阻遏蛋白对于自身基因 *cI* 的正调控作用是一种反馈放大。在生命有机体中的任何正反馈都必须受到控制，否则就会引起生命系统的崩溃。当 λ 阻遏蛋白浓度过高时，就会进入 $O_{R3}$ 位置，从而阻遏了 *cI* 基因的转录，这时 λ 阻遏蛋白又起负调控作用了。在实际上，这一负反馈机制很少使用，也就是说 λ 阻遏蛋白的浓度很少达到能进入 $O_{R3}$ 的地步，这里面可能还有别的因素在起作用。λ 阻遏蛋白对于 $P_R$ 的阻遏作用，对于 $P_M$ 的激活作用和阻遏作用以及 λ 阻遏蛋白二聚体之间的协同效应均示于图 10-6。

由于 Cro 二聚体与 O 位点结合力的顺序恰好与 λ 阻遏蛋白二聚体相反，因此最初表达的 Cro 蛋白二聚体首先与 $O_{R3}$ 结合，从而关闭了 *cI* 基因的活性。这样，由 RecA 切断阻遏蛋白单体分子到阻遏蛋白的失活而引起 Cro 蛋白的合成，再由 Cro 关闭了 *cI* 基因。这些过程不可逆地导致了噬菌体的裂解生长(图 10-8)。

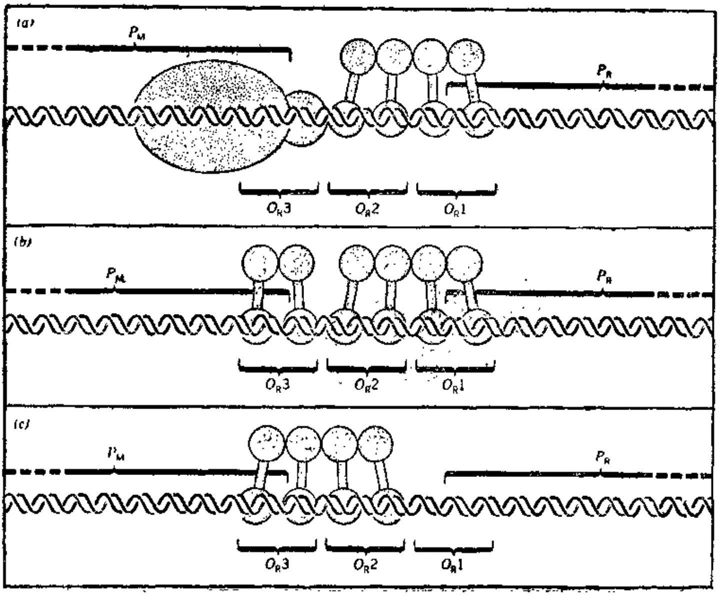
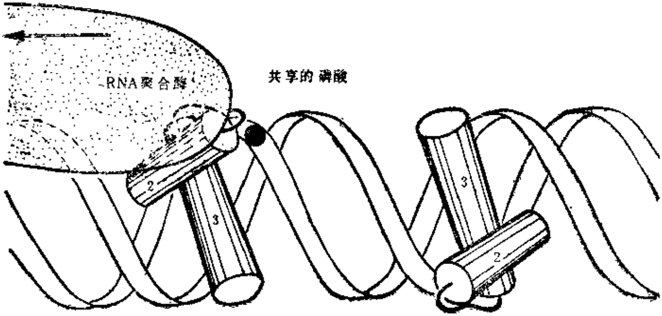



图 10-6 λ 阻遏蛋白二聚体之间以及 λ 阻遏蛋白二聚体与 RNA 聚合酶之间的相互作用控制着右向早期启动子 $P_R$ 和维持溶原状态所必需的 *cI* 基因启动子 $P_M$。

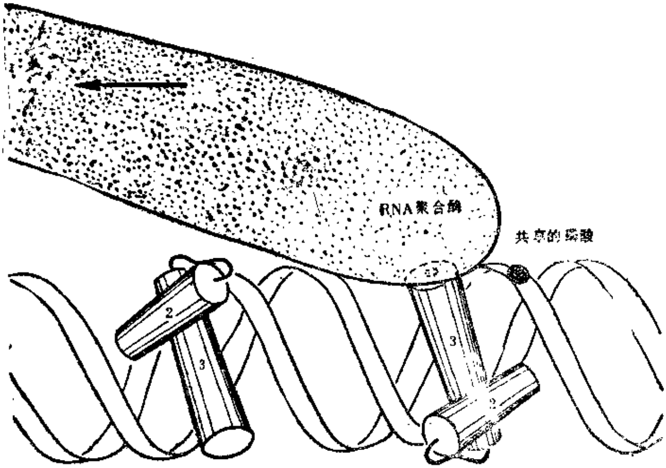
(a) 在通常的溶原状态下，二聚体占据 $O_{R1}$ 和 $O_{R2}$ 阻断了从 $P_R$ 开始的转录，而促进了从 $P_M$ 的转录。

(b) 二聚体协同地占据了 $O_{R1}$ 和 $O_{R2}$，另一个二聚体独立地占据了 $O_{R3}$，从 $P_R$ 和 $P_M$ 的转录都被阻断。

(c) 如果 $O_{R1}$ 位点发生突变，则二聚体可以协同占据 $O_{R2}$ 和 $O_{R3}$，阻断从 $P_M$ 的转录，而从 $P_R$ 的转录也因突变而不能进行。




λ阻遏物




P 22的阻遏物

图 10-7 阻遏蛋白的正调控作用

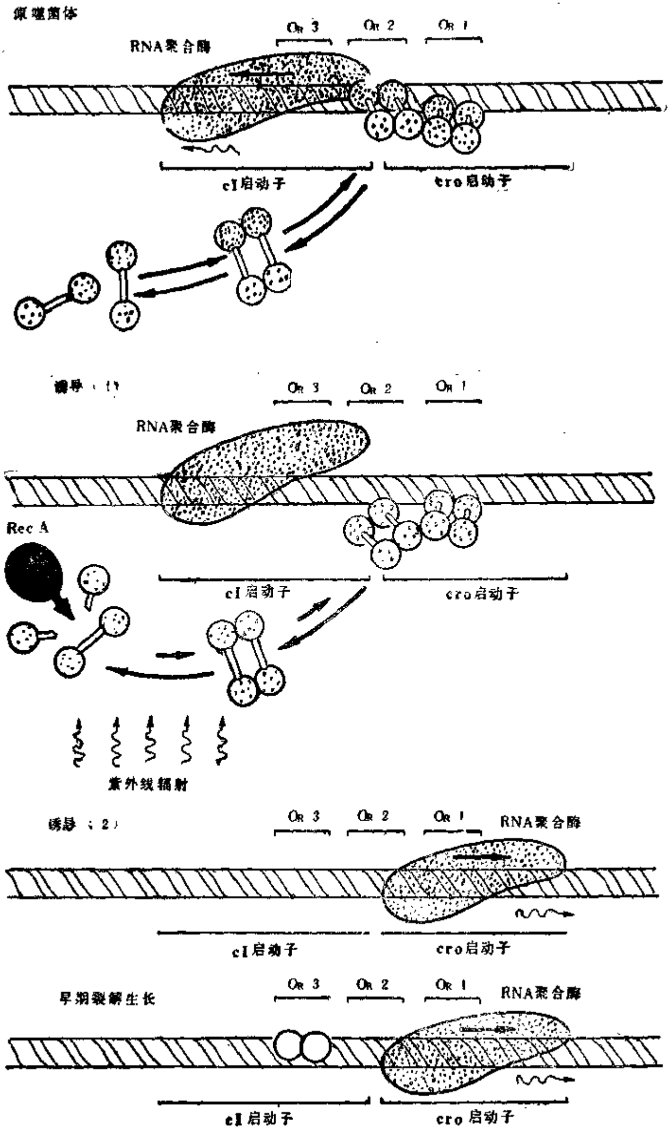



图 10-8　RecA 切断 λ 阻遏蛋白单体，Cro 的合成，*cI* 基因的关闭不可逆地导致了噬菌体的裂解生长。

# 第五节　调控蛋白质的结构花式及其与 DNA 相互作用的细微机制[7~30]

## 一、概　述

调控蛋白质包括负调控因子(阻遏蛋白)和正调控因子(转录因子)。原核生物的启动子(及操作子)的结构较为简单,所涉及的调控蛋白质的种类也较少；真核生物的启动子结构则复杂得多,有的还有增强子结构,因而涉及的调控蛋白质的种类也较多,其中主要是各种转录因子。一个真核生物基因对于环境和发育过程中的不同信号以及这些信号的不同组合将作出不同的响应,即有着不同程度的表达。这一过程的实现,就需要不同的转录因子与特定的 DNA 序列结合,转录因子之间的相互作用以及转录因子与常规转录机构的相互作用。

迄今发现的为数众多的调节蛋白质中，其 DNA 结合结构域大多在 100 个氨基酸以下。这些氨基酸序列脱离整个分子而单独存在时，通常对特异的 DNA 序列也具有相当的结合能

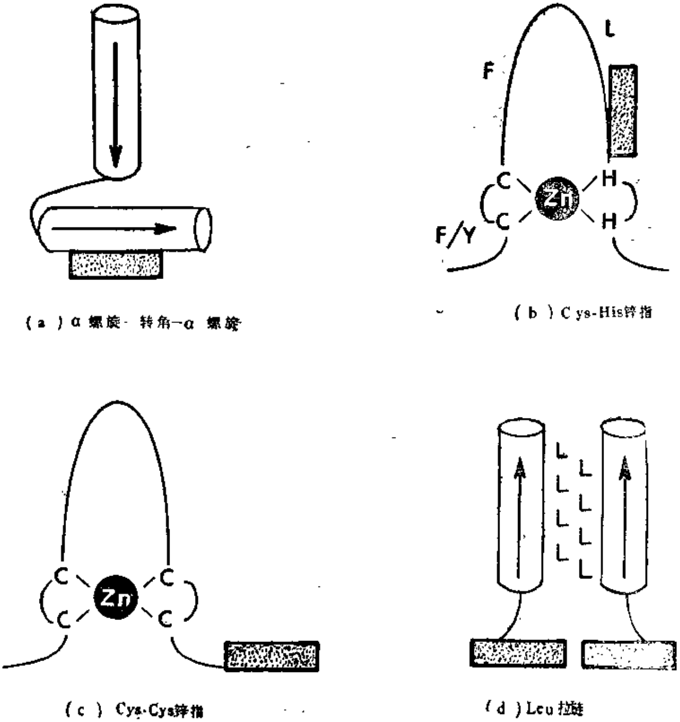



图 10-9　调控蛋白质的 DNA 结合结构域的四种花式的模式图。
带箭头的园柱体表示α 螺旋，带点的框表示 DNA 结合部位。

力。这些分子的 DNA 结合结构域大体上有四种花式：α 螺旋—转角—α 螺旋；Cys—His 锌指；Cys—Cys 锌指；Leu 拉链。这些结构花式对于维持整个结构域的结构是十分重要的，但不是全部参与 DNA 的直接接触。在上述四种花式中，只有第一种花式的作用机制是清楚的，其第二个 α 螺旋(识别螺旋)上的特定氨基酸残基与 DNA 大沟中的碱基发生特异的相互作用。而其余的三种花式的作用机制仍有待于进一步的阐明。现将这四种结构花式的模式图示于图 10-9。

## 二、α 螺旋—转角—α 螺旋

上面我们讲述了 λ 噬菌体的阻遏蛋白和 Cro 蛋白所组成的开关如何控制噬菌体的溶原状态和裂解周期的原理。λ 阻遏蛋白和 Cro 蛋白到底怎样才能选择性地与操作子位点结合而不与 DNA 的其它序列结合呢？所有的调节蛋白质是否使用同一种方案识别其特定的

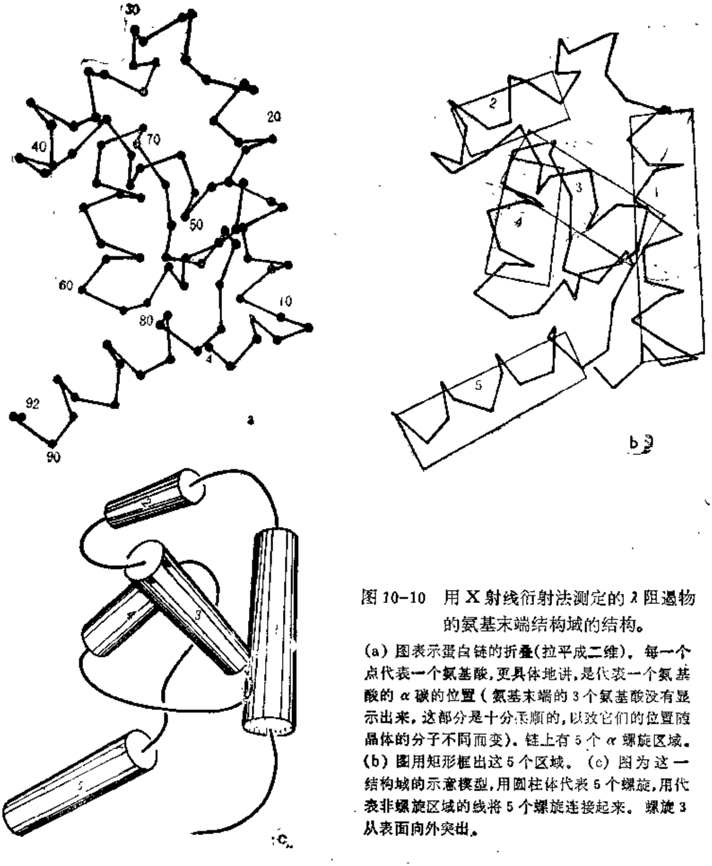


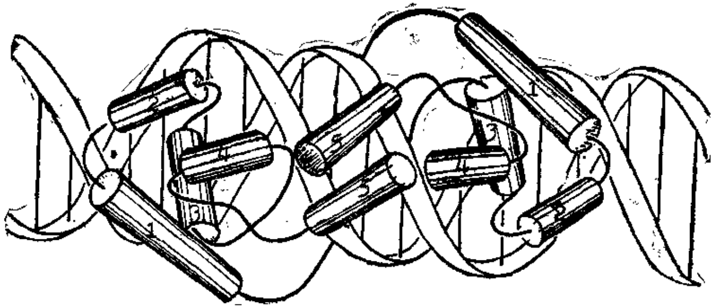

图 10-10 用 X 射线衍射法测定的 λ 阻遏物的氨基末端结构域的结构。

(a) 图表示蛋白链的折叠(拉平成二维)。每一个点代表一个氨基酸，更具体地讲，是代表一个氨基酸的 α 碳的位置(氨基末端的 3 个氨基酸没有显示出来，这部分是十分柔顺的，以致它们的位置随晶体的分子不同而变)。链上有 5 个 α 螺旋区域。(b) 图用矩形框出这 5 个区域。(c) 图为这一结构域的示意模型，用圆柱体代表 5 个螺旋，用代表非螺旋区域的线将 5 个螺旋连接起来。螺旋 3 从表面向外突出。

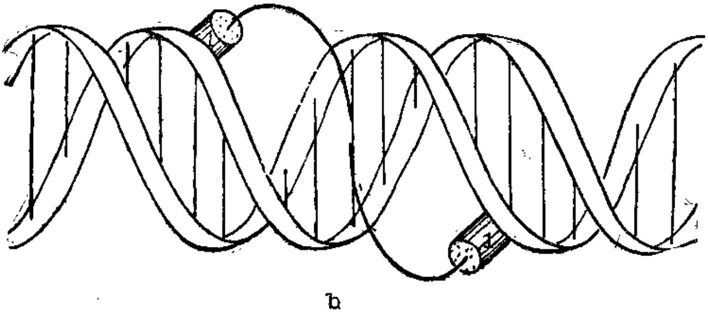
DNA 序列呢？阻遏蛋白调控的本质又是什么呢？要回答这些问题，必须首先了解 λ 阻遏蛋白和 Cro 蛋白的三维结构。

哈佛大学的 Pabo 等人用木瓜蛋白酶切断 λ 阻遏蛋白的两个结构域之间的连接部位，然后将 N- 末端结构域进一步纯化与结晶，再进行 X 光衍射分析。结果发现 N- 末端结构域包含 5 个 α 螺旋由四段转角连接起来；氨基末端有一段长长的和柔顺的臂从主体中伸出来（图 10-10）。

Pabo 等人在弄清 λ 阻遏蛋白的三维结构后，便提出了 λ 阻遏蛋白二聚体与操作子结合的假设的模型(图 10-11)。在这个模型中，α 螺旋 3 嵌入 DNA 的大沟，与碱基对边缘形成特异的相互作用；α 螺旋 2 也接近 DNA，但与 DNA 骨架接触；其余的 α 螺旋不直接与 DNA 作用，但两个阻遏蛋白分子的两个 α 螺旋 5 也可能相互作用以增加复合物的稳定性；整个二聚体的主体位于 DNA 双螺旋的同一侧(前面)，只有氨基末端的两个手臂伸到 DNA 的背后，同样在大沟中与碱基接触。



a



b

图 10-11 λ 阻遏蛋白二聚体与操作子的结合模型

(a) 二聚体的主体部分结合在 DNA 的一侧；

(b) 氨基末端的两个手臂伸到 DNA 的另一侧。

Pabo 等人做出 λ 阻遏蛋白 N 末端结构域的三维结构之前，Matthew 等人就做出 Cro 蛋白的三维结构(图 10-12)。Cro 蛋白的三维结构与 λ 阻遏蛋白不同，它没有氨基末端手臂，有三个 α 螺旋结构，还有一个 β 折叠结构。其中 α 螺旋 2 与 α 螺旋 3 的关系与 λ 阻遏蛋白的 α 螺旋 2 和 α 螺旋 3 的关系非常相似。通过甲基化试验测定受 Cro 蛋白保护的嘌呤碱基，通过乙基化试验测定受 Cro 蛋白保护的磷酸基团，试验结果如图 10-13 所示。受保护的碱基均位于 DNA 双螺旋同一侧的两段大沟中，且呈二倍旋转对称分布。受保护的磷酸基团均位于同一侧面的两段大沟之间。这一试验结果与图 10-91的模型也很相似的。1986 年，ptashne 等人的工作证明了 λ 阻遏蛋白和 Cro 蛋白的 α 螺旋 3 的第一个和第二个氨基酸残

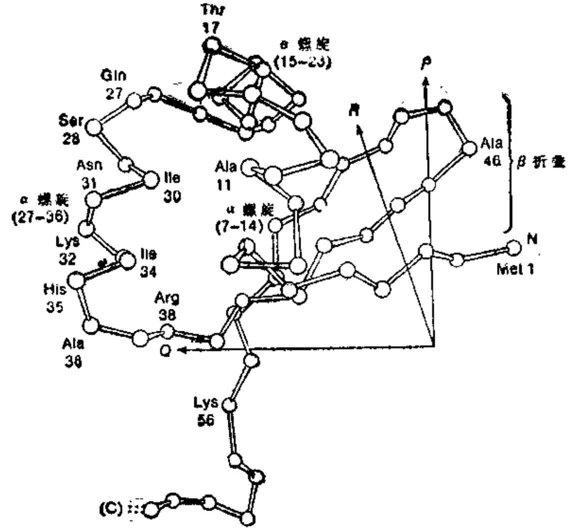



图 10-12 根据晶体 X 光衍射确定的 Cro 蛋白的三维结构

5' T A T C A C C G C A A G G G A T A 3'
3' A T A G T G G C G T T C C C T A T 5'

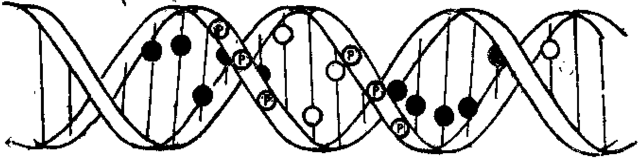



图 10-13 受 Cro 蛋白保护的 $O_{R3}$ 中的嘌呤碱基(黑色圆)和磷酸基团(P)

基(Gln,Ser)是与图 10-13 中第二碱基对的 A 和第四碱基对的 G 相结合的。在 λ 的六个 O 位点中只有这两个碱基是全保守的。而 λ 阻遏蛋白和 Cro 蛋白对于 O 位点的结合力顺序则主要是由第五个和第六个氨基酸残基所决定的。

其他许多调节蛋白质也都有类似于 λ 阻遏蛋白和 Cro 蛋白中 α 螺旋 2 和 α 螺旋 3 的结构,例如耶鲁大学 Steitz 等人测出 CAP 是以其 C 末端结构域与 DNA 作用。Matthew 和 MIT 的 Sauer 等人进一步指出许多调控蛋白与 DNA 特定位点结合的氨基酸序列的保守性,这种保守性均出现在相应于 λ 阻遏蛋白的 α 螺旋 2 和 α 螺旋 3 的这段序列中 (表 10-1)。然后人们普遍认为,调控蛋白质识别 DNA 所采用的 α 螺旋—转角 —α 螺旋的花式是具有代表性的,可能具有普遍的意义(图 10-14)。其中位于大沟中的 α 螺旋对于识别 DNA 具有决定性的意义,称为识别螺旋。λ 阻遏蛋白与操作子的结合情况可见封面图。

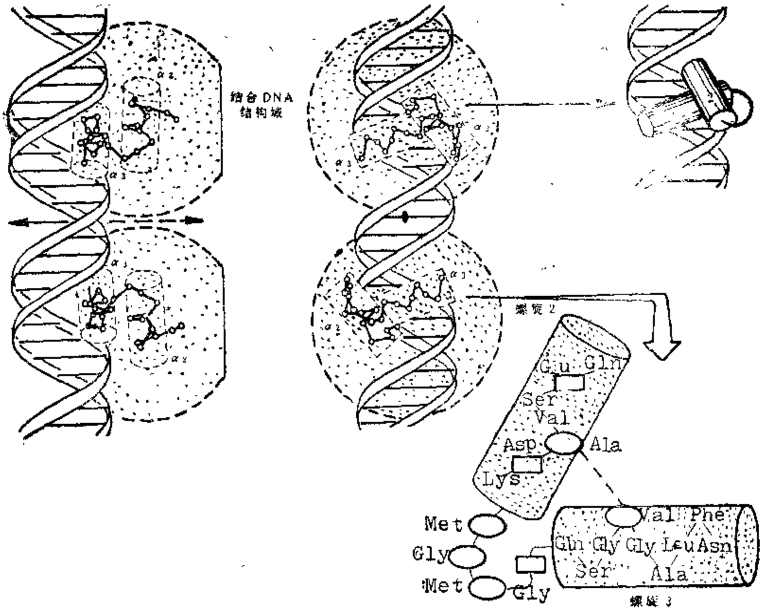



图 10-14 调控蛋白质识别 DNA 所采用的 α 螺旋—转角—α螺旋花式;α螺旋3嵌于大沟中,具有各自的碱基特异性;α 螺旋 2 与磷酸骨架相接触,没有碱基特异性(本图所示为 λ 阻遏蛋白与 $O_{R1}$ 的作用)。放大的两个α螺旋之间由于两个疏水氨基酸之间的相互作用而保持一定的角度。

近几年来,为了证明识别螺旋与 DNA 碱基序列的相互作用,哈佛大学的 Wharton 和 Ptashne做了一系列出色的试验,堪称分子生物学上的换头术。首先把噬菌体 434 的阻遏蛋白中的识别螺旋换成 434 Cro 蛋白的识别螺旋。这种带有434Cro 蛋白识别螺旋的杂种 434 阻遏蛋白分子对于 434 操作子的结合行为完全像434 Cro 蛋白。这可以用化学探针的方法在体外测定出来。这是同种噬菌体的阻遏蛋白和 Cro 蛋白的相互杂交。他们为了进一步确定识别螺

**表 10-1　调控蛋白质的 α 螺旋—转角—α 螺旋单元中的氨基酸序列，带底线的部分为保守的氨基酸，已经做出空间结构的前三种蛋白质的 α 螺旋—转角—α 螺旋结构画于表下面。**

（λ 阻遏蛋白的氨基酸残基序号）

| | 33 | 34 | 35 | 36 | 37 | 38 | 39 | 40 | 41 | 42 | 43 | 44 | 45 | 46 | 47 | 48 | 49 | 50 | 51 | 52 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| λ阻遏蛋白 | GLN | GLU | SER | VAL | ALA | ASP | LYS | MET | GLY | MET | GLY | GLN | SER | GLY | VAL | GLY | ALA | LEU | PHE | ASN |
| λ Cro | GLN | THR | LYS | THR | ALA | LYS | ASP | LEU | GLY | VAL | TYR | GLN | SER | ALA | ILE | ASN | LYS | ALA | ILE | HIS |
| CAP | ARG | GLN | GLU | ILE | GLY | GLN | ILE | VAL | GLY | CYS | SER | ARG | GLU | THR | VAL | GLY | ARG | ILE | LEU | LYS |
| 434 阻遏蛋白 | GLN | ALA | GLU | LEU | ALA | GLN | LYS | VAL | GLY | THR | THR | GLN | GLN | SER | ILE | GLU | GLN | LEU | GLU | ASN |
| 434 Cro | GLN | THR | GLU | LEU | ALA | THR | LYS | ALA | GLY | VAL | LYS | GLN | GLN | SER | ILE | GLN | LEU | ILE | GLU | ALA |
| P22 阻遏蛋白 | GLN | ALA | ALA | LEU | GLY | LYS | MET | VAL | GLY | VAL | SER | ASN | VAL | ALA | ILE | SER | GLN | TRP | GLU | ARG |
| P22 Cro | GLN | ARG | ALA | VAL | ALA | LYS | ALA | LEU | GLY | ILE | SER | ASP | ALA | ALA | VAL | SER | GLN | TRP | LYS | GLU |
| 大肠杆菌 lac 阻遏蛋白 | LEU | TYR | ASP | VAL | ALA | GLU | TYR | ALA | GLY | VAL | SER | TYR | GLN | THR | VAL | SER | ARG | VAL | VAL | ASN |

α 螺旋 2 —— 转角 —— α 螺旋 3

旋是调控蛋白质分子中负责结合专一性的主要决定因素这一假说，他们将大肠杆菌噬菌体434的阻遏蛋白的识别螺旋与沙门氏菌噬菌体 P22 的阻遏蛋白的识别螺旋进行互换，结果带有 P22 阻遏蛋白的识别螺旋的杂种 434 阻遏蛋白在体外不再识别 434 操作子而是识别 P22 的操作子。当时他们并不知道 434 阻遏蛋白和 P22 阻遏蛋白的三维结构，完全是依据表 10-1 中所列的氨基酸序列同系性进行的推理，然而实验却证实了他们的推理。可是上述的杂交分子在细菌体内的试验却不成功。他们进一步进行推理分析后认为，识别螺旋不但要与 DNA 发生作用，而且要与蛋白质分子的其他部分发生作用。整段识别螺旋的替换必然造成识别螺旋与蛋白质分子其余部分的不和谐。识别螺旋与 DNA 作用的一面为外侧面，与蛋白质分子其余部分作用的为内侧面(图 10-15)。如果让内侧面与原来分子保持和谐，而只更换外侧面的氨基酸，则有可能在细菌体内改变该调控蛋白的识别位点。于是，他们将 434 阻遏蛋白的识别螺旋的外侧面换成 P22 阻遏蛋白识别螺旋外侧面的氨基酸。这样的杂种分子称为 434R[α3(P22R)]，见图 10-15。将杂交分子的基因克隆到细胞中，这种细胞不能为 P22 感染所杀死，

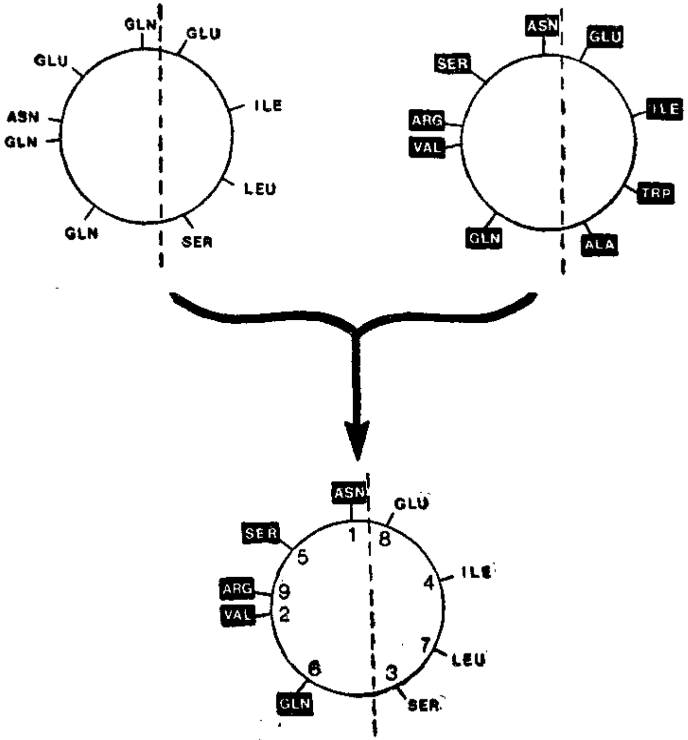



图 10-15 左上图，右上图和下图分别表示 434 阻遏蛋白，P22 阻遏蛋白和杂种阻遏蛋白 434R[α3(P22R)]的识别螺旋。虚线的左面和右面分别表示识别螺旋的外侧面和内侧面。氨基酸编号以 α 螺旋 3 的氨基末端第一个氨基酸为 1 号(参考表 10-1)。

反倒为 434 感染所杀死。(在正常情况下，P22 和 434 为异免疫噬菌体)。这是因为杂交分子 434R[α3(P22R)]失去了识别 434 操作子的能力，因而能够被 434 感染；然而它能够识别 P22 操作子，因此不被 P22 感染。P22 有 6 个为其阻遏蛋白作用的操作子，每一个都有 18bp 长，其通用序列为 ANTNAAGNNNNCTTNANT。434 同样有 6 个操作子，每个都有 14bp 长，通用序列为 ACAANNNNNNTTGT。体外结合试验也表明这两种噬菌体互为异免疫噬菌体，即两种噬菌体的阻遏蛋白和操作子没有交叉反应。而杂交分子 434R[α3(P22R)] 与 P22 操作子具有强烈的结合能力，在 $10^{-9}\sim10^{-8}$mol/L 的浓度下，就能占据一半以上的 P22 操作子。

由以上的各种识别螺旋的交换实验的结果，我们可以看到，识别螺旋的外侧面是它与 DNA 结合专一性的主要的决定因素，434 阻遏蛋白，P22 阻遏蛋白和杂种阻遏蛋白依赖其识别螺旋的外侧面的五个氨基酸残基就足以识别其操作子位点。而阻遏蛋白分子的其余部分很可能是结合所必需的，但对结合的专一性没有贡献；如果说有贡献的话，那也是识别 434 操作子和 P22 操作子的共同特征。434 和 P22 阻遏蛋白的识别螺旋可以互换，说明这些分子将自己的识别螺旋塞入 DNA 大沟中的方式和方向都是很类似的。否则杂种分子 434R[α3 (P22R)]就不可能识别 P22 的操作子。然而，根据已知三维结构的 λ 阻遏蛋白，λ Cro 蛋白和 *E.coli* 的 CAP 蛋白分子模型来看，识别螺旋与 DNA 双螺旋轴的角度有很大的差异，这可能是因为分子模型是根据蛋白质结晶的 X 光衍射分析得出来的。而体内蛋白质并不是结晶状态。细胞内的蛋白质和 DNA 都具有很大灵活性，能够在一定程度上改变自己的构象以相互适应。

在真核生物的转录调控蛋白质中，同样存在着 α 螺旋—转角—α 螺旋结构。首先被人们认为含有这种结构的是酵母接合型调控蛋白质 a1 和 α2，它们是具有序列专一性的 DNA 结合蛋白。随后人们将典型的同形异位蛋白质与 a1 和 α2 进行序列比较，发现相互之间具有高度的序列同源性。许多哺乳动物的启动子和增强子序列中都有八核苷酸序列，其结合的转录因子(Oct1, Oct2)也含有同形异位框，脑垂体中的转录因子 Pit 也同样如此。目前已知具有同形异位框的真核生物蛋白质已有近百种。通过用三维核磁共振法，人们已经证明了同形异位结构域中 α 螺旋—转角—α 螺旋结构的存在。

经典的 Antp 的同形异位结构域的氨基酸序列以及这些氨基酸残基所构成的 α 螺旋 2—转角—α 螺旋 3 (识别螺旋)示于图 10-16。非洲爪蟾的 XlHbox2 和鼠的 Hox2.3 与 Antp 之间只有一个氨基酸残基的差异。通过螺旋交换实验证明，Antp 及其类似的成员，虽然与原核生物具有同样的 α 螺旋—转角—α 螺旋花式，但决定其细微碱基特异性的不是第五和第六个氨基酸残基，而是第九个氨基酸残基。由此可见，真核生物调节蛋白质的 α 螺旋—转角—α 螺旋结构与 DNA 结合的空间关系可能不同于原核生物。

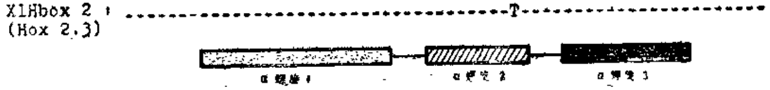



图 10-16 经典的同形异位蛋白质 Antp 的同形异位结构域的氨基酸序列及其 α 螺旋—转角—α 螺旋结构。只有一个氨基酸残基差异的非洲爪蟾 XlHbox2 和鼠 Hox2.3 可能具有相同的 α 螺旋—转角—α 螺旋结构。带星号的为各种同形异位结构域中几乎全保守的氨基酸残基。

## 三、锌指结构

负责转录 5S rRNA 基因，tRNA 基因 Alu 序列和部分 snRNA 基因的 RNA 聚合酶Ⅲ需要三个转录因子：TFⅢA，TFⅢC，TFⅢB。TFⅢA 非同寻常的结构导致了人们对另一类调控蛋白质的 DNA 结合结构域结构花式的建议。

TFⅢA 由 344 个氨基酸残基所组成，分子量为 40 KDa。其 C 端结构域（第 276～344 氨基酸残基）是促进转录活性所不可缺少的。N 端结构域（第 12～276 氨基酸残基）负责与DNA 的结合；在 5SrRNA 基因上结合大约 50bp。N 端的第 1～12 氨基酸残基的功能还不清楚（图 10-17）。TFⅢA 的活性需要 $Zn^{++}$。

图 10-17 TFⅢA 的功能结构域

在 TFⅢA 的 DNA 结合结构域中，存在着九个有规律的正向重复单位。每个单位大约 30 个氨基酸，其中部 10 个氨基酸都是碱性或极性氨基酸。九个重复单位构成了结合 DNA 的九个"手指"。九个单位是一个家族，由同一祖先而来。对 TFⅢA 基因的序列分析证实了这一预言。该基因的 8 个内元中有 7 个是在 DNA 结合手指之间。现在已经发现 20 多种DNA 结合蛋白质都具有锌指结构。典型的锌指蛋白为 $C_2/H_2$ 型，即 $Zn^{++}$ 离子与两个半胱氨基酸残基和两个组氨基酸残基形成配位键。除了这四个氨基酸残基外，还有两个芳香族氨基酸残基和一个亮氨基酸残基也高度保守（表 10-2）。配位结构加上 Phe 和 Leu 所形的疏水核心，确定了锌指的基本结构和形状（图 10-18a）。然而，Berg 等人根据锌指结构中氨基酸残基的亲水/疏水性以及所带电荷，并与其它金属蛋白质进行比较，认为锌指的三维结构可能包含一个β-折叠和一个α-螺旋（图 10-18b）。

表 10-2 几种蛋白质的锌指结构中的保守氨基酸及其位置

| 蛋白质 | 氨基酸序列 | 氨基酸残基数 |
|---|---|---|
| 爪蟾 TFⅢA | 1 3 6 10 16 19 23 29<br>C····C···F·····L··H···H·····Y | 30 |
| 果蝇 Kr | ·C··C···F·····L··H···H·····Y | 28 |
| 果蝇 Sry | ·C··C···F·····L··H···H·····Y· | 29 |
| 酵母 ADR1 | ·C··C···F·····L··H···H·····Y | 28 |

TFⅢA 的九个手指与 5SrRNA 基因内启动子的 50bp 结合，平均每个手指结合 $5^1/_2$bp。Klug 等人对这些结合位点的序列进行分析，发现 G（或 GG，GGG）周期性地出现，相隔约 $5^1/_2$bp。然而很可能这些手指并非平均分配，很可能成簇地与启动子的三个重要区域（OCTGG）相结合。能够与 DNA 结合的可能就是每个手指上的α-螺旋。然而，某些蛋白质

时锌指结构单独存在时并不能与靶 DNA 结合。这就说明锌指结构对于维持 DNA 结构域的结构是必需的,但并不总是与 DNA 直接相互作用。

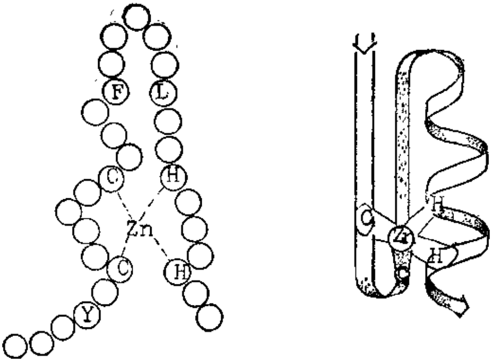



图 10-18 非洲爪蟾 TFⅢA 锌指结构:
(a) 示意图,Zn 与 $C_2/H_2$ 形成配位键,F 和 L 构成疏水核心,图中所示的氨基酸为各种锌指结构中的保守氨基酸。(b)锌指可能存在的三维结构,即 Berg 所建议的包括一个 β- 折叠和一个 α- 螺旋的三维结构。

另一类锌指蛋白为 $C_2/C_2$ 型,其中 $Zn^{++}$ 与四个半胱氨酸残基形成配位键。酵母的转录激活因子GAL4 和哺乳类的固醇类激素受体可以作为典型代表。在这类蛋白质中,如以 $H_2$ 代替 $C_2$,则失去了与 DNA 的结合能力。手指交换试验表明,所有的 $C_2/C_2$ 型锌指蛋白中的锌指结构尽管对于 DNA 结合是必需的,但对结合的专一性并不重要。例如酵母 PPR1 的锌指可以取代 GAL4 的锌指,所生成的杂合蛋白仍然识别 GAL4 的识别位点。因而负责识别专一性的氨基酸序列必定存在于其他部位。

## 四、Leu 拉链

最近,人们发现了一类 DNA 结合结构域的新花式,这就是 Leu 拉链(leucine zipper)。这类蛋白质的主要代表为酵母的转录激活因子 GCN4,癌蛋白 Jun, Fos, Myc,增强子结合蛋白 C/EBP 等。所有这些蛋白都含有 4 或 5 个亮氨酸残基,精确地相距 7 个氨基酸残基。这样,在 α- 螺旋的某一侧面每两圈就出现一个 Leu,这些 Leu 排成一排。这样,两个蛋白质分子的两个 α- 螺旋之间依赖亮氨酸残基之间的疏水作用力形成一条拉链 (图 10-9d)。与这一模型相一致的是, GCN4, C/EBP, Jun, Fos 等都是以二聚体形式与 DNA 结合的, Fos 和 Jun 还能够形成异二聚体与 DNA 结合。GCN4 的 Leu 拉链区这一肽段单独存在时就能形成稳定的二聚体,其中是 100% 的 α- 螺旋。

尽管 Leu 拉链对于二聚体的形成是十分重要的,但是参与 DNA 直接作用的序列显然不在 Leu 拉链区域。不同的 Leu 拉链蛋白都有一个相对保守的区域。GCN4, Jun, Fos 的靶位点的 DNA 序列十分相似,而这些蛋白质分子中与 Leu 拉链区相邻的一段 30 个氨基酸序列也十分相似。在 GCN4 中,Leu 拉链区加上这 30 个氨基酸序列(共 60 个氨基酸残基)就具

有完全的形成二聚体和特异性的 DNA 结合功能。最近引人注目的发现是，Leu 拉链花式也存在于含有锌指花式和同形异位结构域的蛋白质之中。因此 Leu 拉链很可能在多种需要二聚体形式(同聚或异聚)的 DNA 结合蛋白质中充当独立的二聚体形成区域。

## 第六节　不同结合位点上的蛋白质之间的相互作用[8,9,10,30~40]

### 一、蛋白质相互作用的方式

无论是原核生物中还是真核生物中都存在着这样的现象：结合于特定位点上的调节蛋白质能够影响数十个，数百个甚至数千个 bp 以外的基因的转录活性。例如真核生物中结合于增强子上的蛋白质能够影响很长距离以外的 RNA 聚合酶的转录活性。这样的过程是如何实现的呢？原来，一个转录激活因子至少得具备两个功能部位：一是 DNA 结合结构域，二是转录激活区，二者通常位于蛋白质分子的不同部分。就目前所知，大多数转录激活区都有一段包含酸性氨基酸的保守序列，这一序列所形成的结构(或许是 α- 螺旋)处于蛋白质分子的表面。这一酸性激活区与任何一个 DNA 结合结构域融合时都能自主地起激活作用；而DNA结合结构域只是为它在 DNA 分子上寻找一个立足之地。例如利用 *E.coli* LexA 阻遏蛋白的DNA结合结构域(α 螺旋—转角 —α 螺旋)与不同的转录因子的激活区融合都可以产生有功能的杂合转录因子。其中，两个功能部位之间的距离和方向(N 端或 C 端)对其功能并无重大影响。这种酸性激活区在进化上是高度保守的。例如，酵母的 GAL4 可以在哺乳动物细胞中起作用；哺乳动物的 Jun 蛋白和糖皮质激素受体也可以在酵母细胞中促进转录。即使在原核生物中，例如 λ 阻遏蛋白对于 $P_M$ 的正调控作用，也是需要酸性氨基酸残基的(见第四节)。除了酸性氨基酸类型外，有的转录激活结构域富含谷酰胺(如 SP1，果蝇的触角脚基因簇，双胸基因簇等)或富含脯氨酸残基(如 CTF/NFI 等)。

转录因子上的激活区倒底又是怎样实现其激活转录的功能的呢？目前还没有定论。从转录因子在进化过程中的保守性这方面来看，与之作用的也必定是保守的转录机构的成分。在真核生物中，这种成分很可能就是 RNA 聚合酶 Ⅱ 本身或 TATA 结合因子 TFⅡD 等五种通用转录因子。

结合于 DNA 上不同部位的蛋白质因子又是通过什么方式到达一块儿发生相互作用的呢？目前主要有四种学说：拧转说派(twisters)，滑动说派(sliders)，接力说派(oozers)和噜噗说派(loopers)。编者并不赞成通用模式论，因为生物世界千变万化，采用不同手段达到相同的目的，这在生命现象中是屡见不鲜的。在以上四种假说中，噜噗假说得了最广泛的支持，证据也最多，而且很多原来用其他三种假说可以解释的现象，用噜噗假说也能得到很好的解释。

**1．噜噗假说(looping)**　这种假说认为，位于不同的 DNA 结合位点上的两个蛋白质，可以通过两个结合位点之间的 DNA 形成噜噗(loop)得以相互作用，促进或阻遏基因的转录等过程。其中起作用的是两个结合于 DNA 的蛋白质，而不是噜噗本身。

噜噗假说的最早根据是 Adhya 等人对 gal 操纵元的相距 90bp 以上的两个操作子的相互作用的研究以及 Schleif 等人对于 ara 操纵元中位于 −280 的 ara $O_2$ 的研究。这些操纵元中，远距离的操作子对于基因转录的阻遏机制不可能是直接阻碍 RNA 聚合酶的结合，而可能是通过蛋白质的相互作用而产生的，而两个蛋白质之间的 DNA 形成噜噗结构，提供了两个蛋白质相互作用的条件。整圈数的 DNA 双螺旋的插入一般并不影响这种相互作用；而带有

半圈双螺旋(如 0.5,1.5,2.5 等)的插入则严重妨碍了这种相互作用(见第八章第四节)。后来 Schleif 等人提出更多的证据来支持噜噗假说。

这种主张被进一步推广到真核生物中。SV40 的早期基因启动子含有三段调节转录的序列：靠近转录起点的为 TATA 框；在 −49 至 −103 有三个 21bp 的重复序列(称为中段序列)；在远上游(−107～−250)为 144bp 的增强子序列。每个序列上都结合有特定的蛋白质才能有效地转录。Takahashi 等人在增强子序列与中段序列之间插入 5bp 或 15bp 则严重阻碍了转录，而插入 10bp 和 21bp 则影响很小，在中段序列与 TATA 框之间的插入也会产生类似的结果。这就表明，各个序列上所结合的调节蛋白质可能通过噜噗形式而相互作用。

Ptashne 等人研究 λ 阻遏蛋白与其操作子结合时所发现的协同效应(即 λ 阻遏蛋白在 $O_{R1}$ 上的结合使 $O_{R2}$ 这个弱结合位点的结合能力提高 10 倍)，主要就是由于阻遏蛋白羧基端结构域的相互作用。Ptashne 等人进一步在 $O_{R1}$ 和 $O_{R2}$ 之间插入整圈数的 DNA 双螺旋(5,6,7 圈)和半圈数 DNA 双螺旋 (6.5 圈)。结果插入整圈数的例子存在着与野生型差不多的协同效应 (图 10-19)；而插入 6.5 圈的则失去了协同效应。在图 10-19 中，两个 λ 阻遏蛋白二聚体之间的 DNA 变为平滑的噜噗形状，这是与下述观察相符的：这段 DNA 上，周期性地出现对于 DNAase I 的超敏感性和抗性，这两种位点之间相距 5bp。这与真核生物核小体中引起的 DNA 对 DNAase I 的超敏感性的周期性变化是一致的。不同的是这里的 DNA 并没有包绕在蛋白质上。在真核生物染色体上活跃基因处的 DNAase I 的超敏感性或许就是由于噜噗结构引起的。Ptashne 与 Griffith 合作，用电子显微镜观察到相距 5 圈 DNA 的结合位点上 λ 阻遏蛋白相互作用的图象。

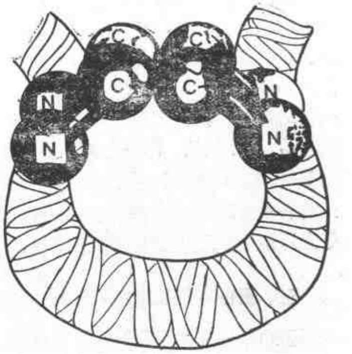


图 10-19　相距 7 圈螺旋的两个 λ 阻遏蛋白结合位点上蛋白质的相互作用

长期被认为只有一个结合位点的 lac 阻遏蛋白，也能以协同的方式结合在两个位点上：一个是经典的操作子位点，叫做 O1；;另一个是位于 *lacZ* 基因内的 O2，与 O1 相距 500bp。调节蛋白质以协同的方式结合于 DNA，增加了 DNA- 蛋白质相互作用的专一性，产生了更为灵敏的遗传开关，噜噗机制使得结合位点相距很远的调节蛋白质能发生相互作用。这就是增强子调控基因转录的基本途径。此外，噜噗机制还提供了基因转录所需的蛋白质 -DNA 复合物的特定拓扑构象。有证据表明，噜噗机制还可能介导了位点特异性重组。

2. *拧转假说*(twisting)　现在大量的实验结果表明，调控蛋白质并不直接结合于显著变形的 DNA，如左旋的 Z-DNA，单链 DNA 等，而是识别具有正常螺旋的 DNA。许多调节蛋白质都像 λ 阻遏蛋白那样，将一个 α 螺旋(识别螺旋)伸到 B-DNA 的大沟中去。曾经一度认为增强子中嘧啶和嘌呤的交替排列所形成的 Z-DNA 与增强子的作用机制是有关的，但是增强子的突变分析表明，这种嘧啶和嘌呤的交替排列 (生成 Z-DNA 所必需) 并不是增强子功能所必需的。关于果蝇中 Z-DNA 与活性基因的关系的实验并不可靠。

然而，有许多实验证据表明，调控蛋白质与其结合位点结合后，是能够改变 DNA 二级结构性质的。例如，cAMP-CAP 的作用。有大量的证据表明，启动子上不与 CAP 结合的DNA

序列的突变能够改变基因转录对于 cAMP-CAP 的依赖性。1987 年 Klug 等人已经证明，cAMP-CAP 与 DNA 结合时，DNA 发生了弯曲。这种弯曲所得到的自由能显著地促进了 RNA 聚合酶结合部位 DNA 的熔解。许多事实表明，当调节蛋白质与 DNA 结合时，DNA 的局部变形是很普遍的现象。通过对类 λ 噬菌体 434 阻遏蛋白与其操作子 DNA 共晶的研究，发现识别螺旋的正确定位需要操作子 DNA 的轻微弯曲和超缠，结果使小沟变狭。而且不直接与蛋白质接触的碱基可以对这种变形作用产生很大的影响；如果将操作子中心部分的几个 A·T 对变为 C·G 对，则使阻遏蛋白与操作子的亲和力降低 40 倍。这可能是由于这样的替代突变削弱了 DNA 的变形适应性。

3．*滑动假说*　即一种调节蛋白质结合于 DNA 的特定序列以后，就沿着 DNA 滑动到另一个特定序列上并与那里的蛋白质发生相互作用从而促进转录的起始。另一种滑动模式是调节蛋白质仍然在原来的位点上，但能用某种机制将 DNA 抽过来，让两种蛋白质发生相互作用。滑动假说最初来源于位点特异性重组现象的解释。但是目前噜噗假说则能更好地解释位点特异性重组。况且，蛋白质离开了它亲和性高的位点向亲和性低的部分滑动，所需的能量也是相当可观的。

4．*接力假说*　即一个调节蛋白质结合于它的结合位点，然后促进了另一种蛋白质结合于相邻的 DNA 序列上，由此再促进别的蛋白质的结合。最后一直伸展到转录起始位点，从而产生对转录的调控。在短距离内，这种作用不是不可能的；然而，在远距离的调控蛋白质（如增强子上），似乎是无法采用这一机制的。

## 三、蛋白质相互作用的复杂性和正、负调控作用的相对性。

蛋白质相互作用包括调控蛋白质之间和调控蛋白质与 RNA 聚合酶之间的相互作用。这些作用机制极为复杂，除了调控蛋白质与 RNA 聚合酶结合位点相互重叠而产生的负调控作用（阻遏）之外，其他作用机制还很不清楚。对于经典的 Lac 阻遏蛋白与 RNA 聚合酶的相互作用机制，人们仍然在不断地完善这方面的认识。起初，人们以为二者在 DNA 上的结合位点相互重叠，似乎是阻遏蛋白的结合使得RNA聚合酶无立足之地。后来通过脚印分析，发现二者在 DNA 上的结合位点并不重叠。人们据此推测可能是由于两个蛋白质分子都很大，阻遏蛋白的结合造成了 RNA 聚合酶结合的空间障碍。上述两种解释的共同点是，阻遏蛋白与 RNA 聚合酶是不能同时与一个 *lac* DNA 相结合的，也就是说二者的结合是相互排斥的。然而，人们新近的研究表明，二者同时结合是可能的。人们用 *lac UV*5 DNA 与 *lac* 阻遏蛋白进行的离体转录试验表明，*lac* 阻遏蛋白对于第一轮的转录表现出瞬时的激活作用。在有无 *lac* 阻遏蛋白的情况下，第一轮转录的速率相差 100 倍。*lac* 阻遏蛋白很可能是阻碍 RNA 聚合酶与启动子复合物的异构化，而不是阻遏 RNA 聚合酶与启动子的结合。各种不同的阻遏蛋白的作用机制是不同的，有的是阻碍 RNA 聚合酶的结合，有的是阻遏 RNA 聚合酶与启动子复合物的异构化，有的是阻止 RNA 合成的起始，甚至有的是阻碍 RNA 链的延伸（如 *gal* 操纵元的阻遏蛋白）。这就表明，阻遏蛋白的作用并无固定的模式。

人们对于调控蛋白质的正调控作用和负调控作用的相对性的认识最早来自 λ 阻遏蛋白，它一方面是 $p_R$ 和 $p_L$ 的负调控因子，同时又是 $p_M$ 的正调控因子；当 λ 阻遏蛋白含量过高时，它又成为 $p_M$ 的负调控因子了。后来，人们又发现了 cAMP-CAP 不但可以作为正调控因子，而且在许多基因的表达中又可以作为负调控因子，如 hfl 基因的表达。最为微妙的却是上述

的 Lac 阻遏蛋白在执行它阻遏功能的同时又预备好在它离开岗位时对转录进行瞬时的激活。

不但经典的阻遏蛋白和分解代谢物激活蛋白(CAP)的调控作用具有两面性，甚至连RNA 聚合酶本身有时也能成为负调控因子。这种情况同样出现在 *lac* 操纵元中。

1981年，T.P.Malan 等人发现 *lac* 操纵元具有两个启动子。一个是早已为人所知的启动子，叫做 *p*1；另一个叫做 *p*2，位于 *p*1 上游方向，二者相距 22 个碱基对。这样，*p*1 的 −35 序列与 *p*2 的 −10 序列相互重叠，*p*2 的 −35 序列与 CAP 位点相互重叠(图 10-20)。*p*2 与 RNA 聚合酶有相当的结合能力，但很难形成开放性启动子复合物。因此，*p*2 启动子起始转录的能力要比 *p*1 差得多。在 *E.coli* 中，当没有 cAMP 存在时，RNA 聚合酶能够与 *p*2 结合；结合于 *p*2 的 RNA 聚合酶便阻碍启动子 *p*1 上 RNA 聚合酶的结合，即 *p*2 上的 RNA 聚合酶起阻遏蛋白的作用。因此，即使当乳糖存在时，如果没有 cAMP-CAP 复合物与其结合位点相结合，RNA 聚合酶就必然先与 *p*2 结合而不能与 *p*1 结合而进行有效的转录。离体试验表明，当事先把 RNA 聚合酶加到含有 *lac* 启动子(包括 *p*1 和 *p*2)的超螺旋 DNA 中并保温，然后再加入 cAMP-CAP 复合物，则启动子 *p*1 的激活速度很慢($t_{1/2}$ = 40 分钟)。相反，如果事先将 cAMP-CAP 与上述 DNA 保温，然后再加入 RNA 聚合酶，则启动子 *p*1 可在 1 分钟内全部被 RNA 聚合酶所占据。cAMP-CAP 与其 DNA 上结合位点的结合，一方

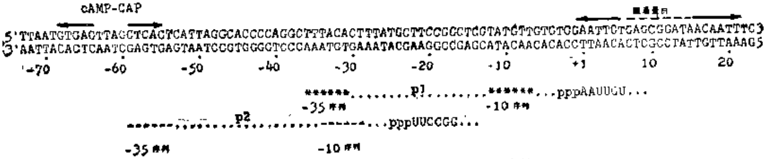



图 10-20 乳糖操纵元的双启动子结构

面成为启动子 *p*2 的阻遏物，另一方面又成为启动子 *p*1 的激活因子。cAMP-CAP 只有阻遏了启动子 *p*2，RNA 聚合酶才有可能与启动子 *p*1 结合。因而，cAMP-CAP 对于 *p*2 的负控制间接地促进了它对于 *p*1 的正控制。cAMP-CAP 对于 *p*1 的正控制，除了以前我们所说过的降低局部双螺旋二级结构的稳定性以促进开放性启动子复合物形成之外，直接与 RNA 聚合酶相互作用也是有可能的。

*lac* 操纵元有两个启动子，又有阻遏蛋白和 cAMP-CAP 的双重调节，因此对于乳糖存在与否是十分敏感的。只有当细胞内缺少葡萄糖而培养基中又有乳糖存在时，*lac* 基因才能有最大的转录活性(图 10-21a)；如果葡萄糖和乳糖都缺乏时，lac 操纵元则达到最大程度的阻遏(图 10-21b)；如果有葡萄糖存在而无乳糖存在时，RNA 聚合酶能够占据 *p*2，以很低频率起

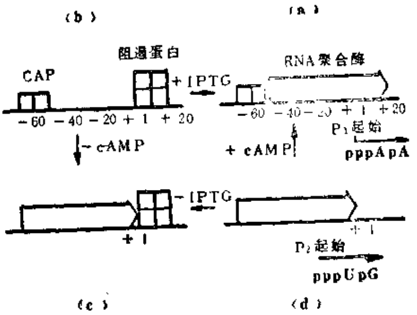



图 10-21 *E.coli lac* 操纵元转录起始调控的模式图。图中以 IPTG 为诱导物。

(a) cAMP-CAP 阻遏了 *p*2 而激活了 *p*1，从而产生最大程度的转录活性；(b) cAMP-CAP 阻遏了 *p*2，阻遏蛋白阻断了 *p*1，实现了最大程度的阻遏；(c) RNA 聚合酶占据了 *p*2，但从 *p*2 起始的转录被阻遏蛋白所阻断；(d) 在缺少 cAMP-CAP 和阻遏蛋白时，能从 *p*2 起始低水平的转录。

始的转录在操作子部位被阻断(图 10-21c);如果葡萄糖和乳糖都存在时,可以从 $p2$ 以很低的频率起始转录而产生 *lac ZYA* 的 mRNA(图10-21d)。即使在图 10-21b 和 c 的情况下,转录也并不是百分之百地被阻断,由于阻遏蛋白的结合与解离的动态平衡,仍然可能存在着频率极低的本底水平的转录。通过启动子的选择来进行转录的调控很可能是 cAMP-CAP所调控的操纵元的普遍特征。半乳糖操纵元也是这样的一个操纵元。

## 第七节 RNA 开关系统[41,42]

RNA 分子上的某一段序列,能够按照不同的条件,有选择性地与两段 RNA 序列之一发生碱基配对,形成局部的二级结构。这种二级结构可以出现在一个 RNA 分子的内部,也可以出现在两个 RNA 分子之间。两种不同的二级结构,代表着基因表达的两种状态或者一个生化过程的前后两个阶段,从而构成一个基因表达的开关系统。RNA 开关系统可以在转录水平起作用,也可以在转录后水平起作用,也可以在翻译水平起作用。为人们所熟知的一个例子就是色氨酸操纵元的衰减子系统,它是通过翻译手段在转录水平上控制 mRNA 合成的终止与否。现在人们已经认识了越来越多的 RNA 开关系统。

在翻译水平上起作用的一个 RNA 开关系统就是抗红霉素基因的 mRNA 中的四段RNA序列能够有选择性地形成两种不同的二级结构(图 10-22)。红霉素能够结合于 23 S rRNA,从而妨碍蛋白质的合成。而抗红霉素基因(由质粒编码)产物是一种甲基化酶,能够特异地将 23 S 核糖体 RNA 上的一个腺苷酸残基甲基化,使红霉素无法结合,而且这种甲基化并不影响核糖体的功能。在甲基化酶编码序列前,同样有一段先导肽编码序列。当细胞内没有红霉素时,核糖体能够顺利地翻译先导肽并从RNA上释放下来,使得序列 1 和 2能够配对,而且不影响序列 3 和 4 的配对,因而核糖体无法与甲基化酶的 S-D 序列结合从而关闭甲基化酶mRNA的

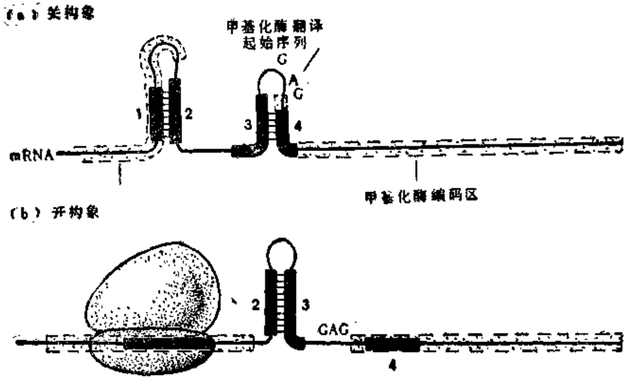



图 10-22 通过先导肽的翻译活动确定红霉素抗性基因 mRNA 中的二级结构状态从而控制甲基化酶翻译的开启与关闭。

(a) 当没有红霉素存在时,核糖体能够顺利地翻译先导肽,使甲基化酶 mRNA 的核糖体结合位点不能与核糖体结合从而阻止其翻译。(b) 当有红霉素存在时,结合有红霉素的核糖体停顿于序列 1,致使序列 2 和 3 配对,从而使 GAG 能被核糖体识别而进行翻译

翻译活性。当细胞内出现红霉素时，结合有红霉素的核糖体便停顿于先导肽的编码区(序列1附近)，从而破坏了序列1和2的二级结构，然后序列3和4的二级结构也受到破坏，而形成序列2和3之间的二级结构（序列2和3形成二级结构所释放的自由能大于序列3和4所释放的自由能)。这样，甲基化酶基因的S-D序列便处于二级结构之外，尚未结合红霉素的核糖体（或者由于甲基化酶本底水平的合成而出现的已甲基化的核糖体）就翻译出甲基化酶。当所有的核糖体都被保护(甲基化)时，先导肽编码区就不再有核糖体停顿，甲基化酶的翻译活动又被关闭。

在转录后处理过程中，U1 snRNA（以RNP的形式)先与5′拼接点附近的序列结合，在5′拼接点被切断后便转向与3′拼接点附近序列结合。这是两个RNA分子所构成的开关系统(见第六章)。这一开关系统的转换机制目前还不清楚。毫无疑义的是，人们将会发现更多的RNA开关系统。

## 提　要

原核生物和真核生物尽管存在着许多差别，基因表达的调控方面也各有各的花样，但二者具有惊人的平行关系和相似性。在从DNA合成，RNA合成直到蛋白质合成这样的遗传信息流动过程中，以转录过程的调控最为重要。

与蛋白质编码信息相比，调控信息的编码方式和解读都存在着自己的特点。调控信息的编码是三维的，具有很高的简并性，调控信息在解读时并不中断原有的Watson-Crick碱基配对，而是调控蛋白质与大沟(或小沟)中的潜在的氢原子供体或受体之间形成氢键，并能引起较大范围的双螺旋结构变化。调控密码的概念实质上就是指调控蛋白质与DNA的相互作用和不同调控蛋白质相互之间以及它们与RNA聚合酶之间的相互作用的规律。

调控蛋白质均为球形蛋白质。为了提高调控的精确性，调控蛋白质通常是同种或异种蛋白质的二聚体或四聚体。同种蛋白质的二聚体或四聚体可能存在着许多种对称性，而邻近的两段DNA序列的对称性则受到限制，只有反向重复序列和正向重复序列两种，其中以前一种居多。

λ噬菌体的阻遏蛋白和Cro蛋白对于左右两个早期操纵元的操作子上的三个结合位点有不同的亲和力，因而构成了控制λ噬菌体的溶原生长和裂解生长的分子开关。二者都是通过α螺旋—转角—α螺旋的结构与DNA结合的。

迄今发现的为数众多的调控蛋白质中，其DNA结合结构域大多在100个氨基酸以下。这些结构域大体上有四种花式：α-螺旋—转角—α-螺旋；Cys—His锌指；Cys—Cys锌指；Leu拉链。这些结构花式对于维持整个结构域的空间结构是十分重要的，但不是全部参与DNA的直接接触。真核生物的同形异位框这一部分也是采用α-螺旋—转角—α-螺旋花式，但是这一结构花式与DNA的空间关系可能不同于原核生物。

转录因子除了DNA结合结构域，还必须有转录激活区。转录激活区主要有酸性氨基酸型，富含谷酰胺型，富含脯氨酸型。这些转录激活区必定是与通用的转录机构的组分相互作用。这些组分很可能就是RNA聚合酶Ⅱ本身或TATA框的结合因子TFⅡD等五种通用转录因子。结合于DNA上不同位点的蛋白质又是通过什么方式到达一块儿发生相互作用的呢？各种机制中，嗜噗假说得到了最为广泛的支持。

由于调节蛋白质之间相互作用的复杂性，因此就产生了正负调控的相对性。人们发现λ

阻遏蛋白一方面是 $P_R$ 和 $P_L$ 的负调控因子，同时又是 $P_M$ 的正调控因子，当 λ 阻遏蛋白含量过高时，它又成为 $P_M$ 的负调控因子。近来人们还发现 *lac* 阻遏蛋白对 RNA 聚合酶的第一轮转录有瞬时的激活作用。甚至是一个结合于启动子邻近位点的 RNA 聚合酶能够成为另一个 RNA 聚合酶与启动子结合的阻遏物。

除了调节蛋白质外，人们还发现 mRNA 的不同二级结构能影响基因的表达。色氨酸操纵元的衰减子就是通过翻译来控制转录终止子二级结构的形成与否。抗红霉素基因 mRNA 的二级结构控制了该 mRNA 的可翻译性。在 RNA 拼接过程中，U1 snRNA 先后与 5′ 拼接点和 3′ 拼接点形成碱基配对，可以看成是两个 RNA 分子组成的开关系统。

## 参 考 文 献


1. Dickerson, R. E. (1983) "The DNA Helix and How It Is Read" *Sci. Amer*. 249: 87～102.
2. Pabo, C. and Sauer, R. (1984) "Protein-DNA Recognition." *Ann. Rev. Biochem*. 53: 293～321.
3. Zubay, G. (1983) *Biochemistry* (Addison-Wesley Publishing Company) pp. 108～109.
4. Robertson, M. (1987) "Gene Regulation: Specificity and Flexibility." *Nature* 327: 464～466.
5. Johnson, A. D. et al. (1981) "λ Repressor and Cro-components of an Efficient Molecular Switch." *Nature* 294: 217～223.
6. Hochschild, A. et al. (1983) "Repressor Structure and the Mechanism of Positive Control." *Cell* 32: 319～325.
7. Wharton, R. and Ptashne, M. (1985) "Changing the Binding Specificity of a Repressor by Redesigning an α-helix." *Nature* 316: 601～605.
8. Bushman, F. D. and Ptashne, M. (1988) "Turning λ Cro into a Transcriptional Activator." *Cell* 54: 191～197.
9. Struhl, K. (1989) "Helix-Turn-Helix, Zinc-Finger, and Leucine-Zipper Motifs for Eukaryotic Transcriptional Regulatory Proteins." *Trends Boichem. Sci*. 14: 137～140.
10. Brennan, R. G. and Matthews, B. W. (1989) "Structural Basis of DNA-Protein Recognition." *Trends Biochem. Sci*. 14: 286～290.
11. Pabo, C. and Lewis, M. (1982) "The Operator-Binding Domain of Lambda Repressor: Structure and DNA Recognition." *Nature* 298: 443～447.
12. Anderson, W. et al. (1982) "Proposed α-helical Super-secondary Structure Associated with Protein: DNA Recognition." *J. Mol. Biol*. 159: 745～751.
13. Sauer, R. et al. (1982) "Homology among DNA-binding Proteins Suggests Use of a Conserved Super Secondary Structure." *Nature* 298: 447～451.
14. Steitz, T. et al. (1982) "Structural Similarity in the DNA-Binding Domains of Catabolite Gene Activator and Cro Repressor Proteins." *Proc. Nat. Acad. Sci*. 79: 3097～3100.
15. Hochschild, A. et al. (1986) "Homologous Interactions of λ Repressor and Cro with the Operator." *Cell* 44: 925～933.
16. Wharton, R. P. and Ptashne, M. (1986) "An α-helix Determines the DNA-binding Specificity of a Repressor." *Trends Biochem. Sci*. 11. 71～73.
17. Wright, C. V. E. et al. (1989) "Vertebrate Homeodomain Proteins: Families of Region-Specific Transcription Factors." *Trends Biochem. Sci*. 14: 52～56.
18. De Robertis, E. M. et al. (1989) "Determination of Axial Polarity in the Vertebrate Embryo: Homeodomain Proteins and Homeogenetic Induction." *Cell* 57: 189～191.
19. Hanes, S. D. and Brent, R. (1989) "DNA Specificity of the Bicoid Activator Protein Is Determined by Homeodomain Recognition Helix Residue 9." *Cell* 57: 1275～1283.
20. Levine, M. and Hoey, T. (1988) "Homeobox Proteins as Sequence-Specific Transcription Factors." *Cell* 55: 537～540.
21. Vincent, A. (1986) TFⅢA and Homologous Genes, the Finger Proteins." *Nucl. Acid Res*. 14: 4385～4391.
22. Frankel, A. D. and Pabo, C. O. (1988) "Fingering Too Many Proteins." *Cell* 53: 675.
23. Berg, J. M. (1988) "Proposed Structure for the Zinc-Binding Domains from Transcription Factor ⅢA and Related Proteins." *Proc. Nat. Acad. Sci*. 85: 99～102.

24. Rhodes, D. and Klug, A. (1986) "An Underlying Repeats in Some Transcriptional Control Sequences Corresponding to Half a Double Helical Turn of DNA." *Cell* 46: 123～132.
25. Pieler, T. et al. (1987) "The 5S Gene Internal Control Region Is Composed of Three Distinct Sequence Elements, Organized As Two Functional Domains with Variable Spacing." *Cell* 48: 91～100.
26. Evans, R. M. and Hollenberg, S. M. (1988) "Zinc Fingers: Gilt by Association." *Cell* 52: 1～3.
27. Landschulz, W. H. et al. (1988) "The Leucine Zipper: A Hypothetical Structure Common to a New Class of DNA Binding Proteins." *Science* 240: 1759～1764.
28. O'Shea, E. K. et al. (1989) "Evidence That the Leucine Zipper Is a Coiled Coil" *Science* 243: 538～542.
29. Sturm, R. A. and Herr, W. (1988) "The TOU Domain Is a Bipartite DNA Binding Structure." *Nature* 336: 601～604.
30. Ptashne, M. (1988) "How Eukaryotic Transcriptional Activators Work." Nature 335: 683～689.
31. Struhl, K. (1988) "The Jun Oncoprotein, a Vertebrate Transcription Factor, Activates Transcription in Yeast." Nature 332: 649～650.
32. Mitchell, P. J. and Tjian, R. (1989) "Transcriptional Regulation in Mammalian Cells by Sequence-Specific DNA Binding Proteins." Science 245: 371～378.
33. Goldberg, R. B. et al. (1989) "Regulation of Gene Expression during Plant Embryogenesis." Cell 56: 149～160.
34. Schleif, R. (1987) "Why Should DNA Loop" *Nature* 327: 369～370.
35. Hochschild, A. and Ptashne, M. (1986) "Cooperative Binding of λ Repressors to Sites Separated by Integral Turns of the DNA Helix." *Cell* 44: 681～687.
36. Ptashne, M. (1986) "Gene Regulation by Proteins Acting nearby and at a distance." Nature 322: 697～701.
37. Travers, A. and Klug, A. (1987) "DNA Wrapping and Writhing." Nature 327: 280～281.
38. Zheng, X. -M. et al. (1987) A General Transcription Factor Forms a Stable Comlpex with RNA Polymerase Ⅱ." *Cell* 50: 361～368.
39. Stramey, S. B. and Crothers, D. M. (1987) "Lac Repressor Is a Transient Gene Activating Protein." *Cell* 51: 699～707.
40. Malan, T. P. and McClure, W. R. (1984) "Dual Promoter Control of the E. coli Lactose Operon." *Cell* 39: 173～180.
41. Horihouchi, S. and Weisblum, B. (1981) "The Control Region for Erythromycin Resistance: Free Energe Changes Related to Induction and Mutation to Constitutive Expression." *Mol. Gen. Genet.* 182: 341～348.
42. Wickens, M. P. and Dahlberg, J. E. (1987) "RNA-protein Interactions." *Cell* 51: 339～342.

# 第十一章 遗传重组

## 第一节 概　述

生物进化需要可遗传的变异不断产生，变化的根本原因是突变，但就每个个体而言发生突变的机率毕竟很低。涉及到的基因数目非常有限，如果只有突变而没有不同个体间的基因交流。则生物体难以迅速地“组装”产生最能适应环境条件的基因组合。而且在积累具有选择优势的突变的同时也不可避免地积累许多难以摆脱的不利突变。而通过不同个体间的基因交流，可以将具有选择优势的和选择负势的基因分开，在短期内产生具有各种基因型的个体，接受自然选择的检验，从而使整个群体维持在一种不断地淘汰有害等位基因和不断地积累有利基因的动态变化之中。

从广义上讲，任何造成基因型变化的基因交流过程都叫遗传重组（genetic recombination)。真核生物减数分裂时，通过非同源染色体的自由组合形成各种不同的配子，雌雄配子结合后产生基因型各不相同的后代。这种重组过程导致基因型的变化。但不涉及 DNA 分子内的断裂-复合，因此不包括在分子遗传学对重组的研究范围之内。

狭义上的遗传重组指涉及到 DNA 分子内断裂-复合的基因交流，有时又叫做交换(crossing over)。根据对 DNA 序列和所需蛋白质因子的要求，可以把重组分为四类，即同源重组(homologous recombination)、位点特异性重组(site-specific recombination)、转座作用(transposition) 和异常重组 (illegitimate recombination)。大肠杆菌中这四种重组的一般特征见表 11-1。

表 11-1　大肠杆菌中四种遗传重组的一般特征

| 类　　型 | 是否需要序列同源性 | 是否需要 RecA蛋白质 | 是否需要序列特异性酶 |
|---|---|---|---|
| 同源重组 | 是 | 是 | 否 |
| 位点特异性重组 | 是 | 否 | 是 |
| 转座作用 | 否 | 否 | 是 |
| 异常重组 | 否 | 否 | ? |

同源重组发生在 DNA 的同源序列之间。真核生物非姊妹染色子体的交换、姊妹染色子体的交换、细菌及某些低等真核生物的转化、细菌的转导、接合、噬菌体的重组等都属于这一类型。同源重组中负责 DNA 配对和重组的蛋白质因子无碱基序列特异性，只要两条 DNA 序列相同或接近相同，就可以在序列中的任何一点发生。当然也存在重组热点，即某类序列发生重组的机率高于其他序列。此外，真核生物中染色质状态对重组也有影响，比如异染色质及附近区域很少发生重组。大肠杆菌的同源重组需要RecA蛋白质，类似的蛋白质也存在于其它细菌中。因

此细菌中的同源重组又叫做依赖 RecA 重组(RecA dependent recombination)。同源重组中的一个特例是相外同源重组，其结果是产生了非均等交换。这一点我们在第二章中已经讲过。

位点特异性重组发生在两条 DNA 的特异位点上。λ 噬菌体 DNA 通过其 *attp* 位点和大肠杆菌 DNA 的 *att B* 之间的位点特异性重组而实现整合过程。在重组部分有一段 15bp 的同源序列，它是重组的必要条件，但不是充分条件，还须有位点特异性的蛋白质因子参与催化。这些蛋白质因子不能催化其他任何两条序列(不管是同源的还是非同源的)之间的重组，这就保持了 λ 噬菌体 DNA 整合方式的特异性和高度保守性。因此，位点特异性重组又常叫做保守重组(conservative recombination)。位点特异性重组不需要 RecA 蛋白质的参与。

转座过程中，转座成分从染色体的一个区段转移到另一个区段或从一条染色体转移到另一条染色体。转座作用既不依靠转座成分和插入区段序列的同源性，又不需要 RecA 蛋白质。由于转座作用总是伴随着转座成分的复制，故又称为复制重组(replicative recombination)。

异常重组也不需要 DNA 序列的同源性和 RecA 蛋白质，而且它也不需要转座酶。目前对这类重组的机制几乎一无所知。

需要指出的是，这里所说的遗传重组和重组 DNA 技术不是一回事，前者指发生在生物体内的基因交流，是我们要研究、要揭开其奥秘的生物学现象；后者指在体外人为地将不同源的 DNA 组合在一起，是我们有目的有计划地改造生物体的一项技术和手段。二者的原理和机制也相差甚远。本书对于重组 DNA 技术不可能作详细的叙述。在第十二章(基因工程导论)中我们将介绍重组 DNA 的基本原理。

## 第二节 同源重组的分子机制[1~6]

### 一、断裂-复合及 Holliday 中间体的形成

同源重组涉及到参与重组的双方 DNA 分子的断裂与交换复合。断裂-复合的最初证据来自细胞遗传学对减数分裂时染色体行为的研究。减数第一次分裂前期配对的两条非姊妹染色体之间发生断裂-复合，出现交叉结构。再通过交叉的歧化完成重组过程。通过特殊的染色体标记技术可以在显微镜下直接验证断裂-复合的发生。

重组的前提之一是参与重组的两条 DNA 分子必须靠近，同源序列之间严格配对，配对又叫做联会(synapsis)。原核生物的 DNA 无蛋白质规律性地包被，配对多在裸 DNA 分子之间进行，目前尚难直接观察到。真核生物的 DNA 组织在染色体中，配对在染色体之间进行，发生在减数分裂的极早期。参与联会的双方实际上是两条已复制的同源染色体，但也有人认为联会和复制同时发生。联会时，非姊妹染色子体之间以侧面紧靠在一起，形成联会丝复合体(synaptonemal complex)。图 11-1 就是这种复合体的照片。

处在联会丝复合体中间是中心成份(central element)，直径约 18nm；两边各有一个侧向成份(lateral element)直径约 50nm，在电镜下呈横向条痕状；再向外才是染色子体。中心成分和侧向成分处于同一平面，平行延伸。非姊妹染色子体之间不直接接触，而是相距 200nm 以上。这一距离从分子水平上说是相当大的，因此光是联会丝复合体的形成仍不能使两条 DNA 分子接近到可以互相作用发生重组的范围之内，还必须有别的因素参与才行。电镜下可以看到联会丝复合体上偶而出现将两条染色子体联在一起的节或瘤(node)，它们出现的频率和分布

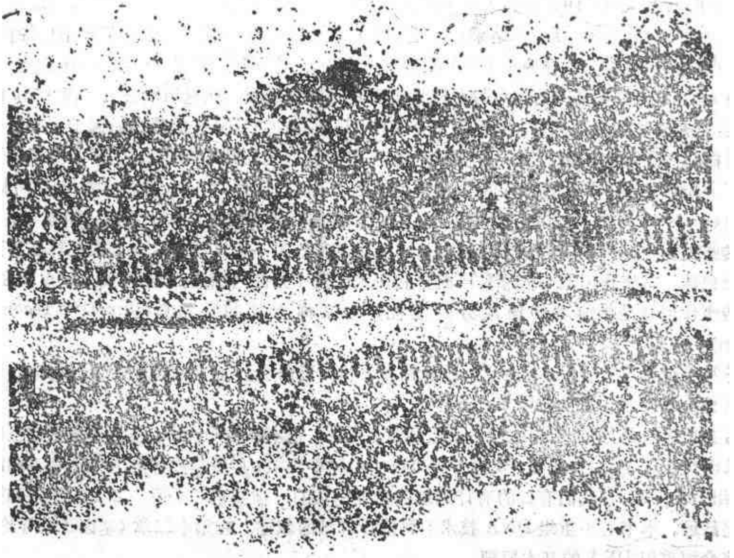


图 11-1　真核生物减数分裂时联会丝复合体照片，ce 代表中心成分，le 为侧向成份，Xt 为染色质

与染色体交叉的频率和分布十分相似，有人认为它们可能就是重组发生的部位。因此又把这些节或瘤叫做重组小瘤(recombination nodule)。联会丝复合体中存在染色体交叉现象，可被光学显微镜观察到，在细胞分裂前夕消失。

目前尚不能对真核生物的遗传重组直接进行分子水平的分析。不过，通过对重组的遗传学后果的分析，对减数分裂早期染色体行为的观察分析，对离体重组系统(主要为原核生物)的分析和对 DNA 生物化学特性的分析，现在已经提出了各种旨在解释双链 DNA 之间重组的可能机制。本节将介绍当前较为流行的模型，这一模型最初由 Holliday 提出，后经 Meselson 和 Radding 修改。图 11-2 为该模型对重组过程的解释，可以分为以下几个步骤：

1．切断(nicking)　同源联会的两个 DNA 分子中任意一个出现单链切口，切口由某些 DNA 内切酶产生。很可能在正常情况下 DNA 分子上随机发生这类切割和随后由连接酶催化的连接。也有可能使同源 DNA 接近并发生联会的蛋白质因子同时具有内切酶活性，负责切口的产生。

2．链置换(strand displacement)　切口处形成的 5′ 端局部解链，由细胞内类似于大肠杆菌DNA聚合酶I的酶系统利用切口处的 3′-OH 合成新链，而把原有的链逐步排挤置换出来，使之成为游离的以5′-P为末端的单链区段，单链置换反应可以一直进行下去，由此产生的单链区段随之越来越长。

3．单链侵入(single-strand invasion)　由链置换产生的单链区段侵入到参与联会的另一条DNA 分子因局部解链而产生的单链泡中。局部解链可能由某种 DNA 结合蛋白质的作用产生，也可能由 DNA 呼吸作用产生。大肠杆菌中，单链侵入需要 RecA 蛋白质的参与。

4. 嗜噗切除(loop cleavge)　侵入的单链 DNA 与参与联会的另一条 DNA 分子中的互补链形成碱基配对，同时把与侵入单链的同源链置换出来，由此产生 D-啥噗。D-啥噗的单链区随后被切除降解掉。这一步骤至少需要一个内切反应和5′→3′外切酶活性。外切作用可能扩展到整个 D-啥噗单链区及其附近区段。

5. 链同化(strand assimilation)　啥噗切除中产生的 3′-OH 断头和侵入单链的 5′-P 由 DNA 连接酶共价连接。同时，侵入的单链可以沿 5′→3′ 方向继续置换出啥噗切除中产生的 5′-P断头，后者不断被 5′→3′ 外切酶活性切除降解。通过单链侵入和链同化作用，被侵入的 DNA 双螺旋分子上有一段区域含有来自联会对方的一条链，它随链同化作用的继续进行而逐步扩大。这一区域叫做杂合DNA (hybrid DNA) 或异源双链(heteroduplex)。因为参与重组的两条同源 DNA 的碱基序列不一定完全一致，异源双链区内往往含有错配碱基，这些错配碱基对面临着细胞内修复系统的修复。异源双链区的形成在立体结构上几乎没有什么困难和障碍，很少造成较长的未配对的碱基区段。到此为止异源双链区只出现在两条 DNA 分子的一条之中，因此这段异源双链区叫做非对称(asymmetric)异源双链区。

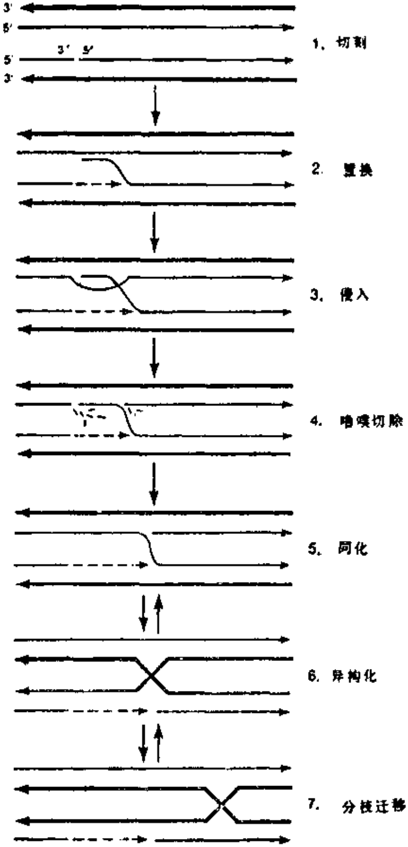


图 11-2　Meselson 和 Radding 提出的关于同源重组机制的链转移模型

6. 异构化(isomerization)　链同化进行过程中，DNA 经过一定的扭曲旋转而形成如图 11-2 第 6 步所示的异构体。这样，两条 DNA 分子不再以一条链相连而是以两条同源的单链交叉相连，而且相连的链不再是前面一系列变化中涉及的单链而是原来一直“安然未动”的链。图 11-2 第 6 步所示的结构最初由 Holliday 提出，故叫做 Holliday 中间体 (Holliday intermediate)，Holliday 中间体的形成机制见图 11-3。其中 I 和 V 所示的结构可以互变，采用分子模型的模拟实验证明图 11-3所示的 DNA 分子异构化过程很容易发生，并不造成任何立体结构上的张力。

7. 分支迁移(branch migration)　两条 DNA 分子之间形成的交叉点可以沿 DNA 移动，这一过程叫做分支迁移。迁移实际上是两条 DNA 分子之间交叉的同源单链互相置换的结果，迁移的方向可以朝向 DNA 分子的任意一端。分支迁移与 DNA 呼吸作用有关，呼吸作用产生的单链区为两条 DNA 分子间的进一步单链取代提供了方便。由于 DNA 呼吸作用此

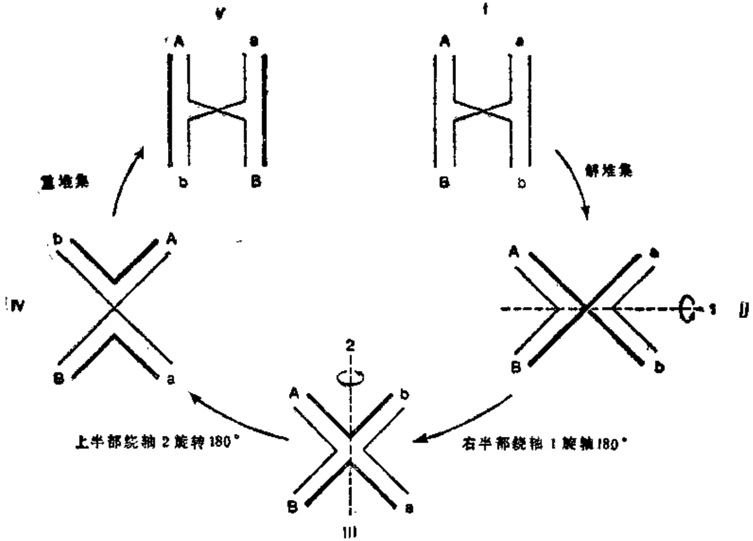



图 11-3　Holliday 中间体的异构化，I 和 V 可以通过异构化而互变

起彼伏，不断发生，所以交叉点也不断处于左右移动之中。长距离的分支迁移需要 DNA 的拓扑异构酶系统的参与。分支迁移使两条 DNA 分子中都出现异源双链区，有时叫做对称(symmetric)异源双链区，其长短视分支迁移情况而异，有时可达数千碱基对。分支迁移改变了交叉点的位置，但 Holliday 中间体的基本结构未变。

Holliday 中间体的形成机制一直是长期争论的问题。除了上述的 Meselson-Radding 模型(又称为链转移模型)外，还有最早由 Whitehouse 提出的 Whitehouse 模型，Hollliday 自己提出的 Holliday 模型(为 Meselson-Radding 模型的原型)以及所谓"先配对后切割"模型("pair, then cut" model)等等。所有这些模型都有相同的基本特征，即作为重组中间体的异源双链区的形成以及使之扩大的分支迁移。近年来大量的实验结果尤其是对大肠杆菌 Rec 重组系统的精细研究多数支持 Meselson-Radding 模型，所以这里我们将不再对其它模型作详细介绍。

## 二、Holliday 中间体的拆分

Holliday 中间体的形成只完成了重组的一半，由它连锁在一起的两条 DNA 分子又必须经过拆分(resolution)回复到彼此分开的双螺旋分子状态。拆分需要内切酶在交叉点处形成一对切口，然后由 DNA 连接酶连接。由于交联在一起的两条 DNA 分子实际上不断处在图 11-3 所阐述的异构变化中，切口可能在两对同源链中的任意一对上发生，如图11-4所示。如果切口发生在形成 Holliday 中间体过程中一种直保持完整的一对同源单链上，至此所有四条单链实际上都已经发生过断裂，由此将释放出两条重组DNA分子，一条亲体双螺旋分子和另一条亲体分子共价相连，中间有一段异源双链区域，这种重组叫做交互重组(reciprocal recombi-

nation)。如果重组分子中的两个遗传学位点位于这段异源双链区的两侧，就可以在表现型上观察到它们之间经常的重组现象，并且重组总是交互的(图 11-4 右)。如果切口发生在重组过程中曾被切开过的一对同源单链上，那么原来一直保持完整的另一对同源链将继续保持完整。这种切割释放出两条原来的亲型 DNA 分子，无交互重组发生，但每条分子上有一段异源双链区。如果两个遗传学位点处在这段异源双链区的两侧，将在表现型上观察不到重组的发生(图 11-4 左)。

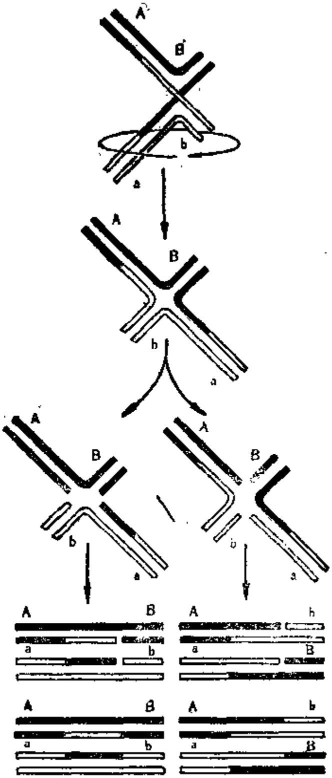



图 11-4　Holliday 中间体拆分的两种可能情况

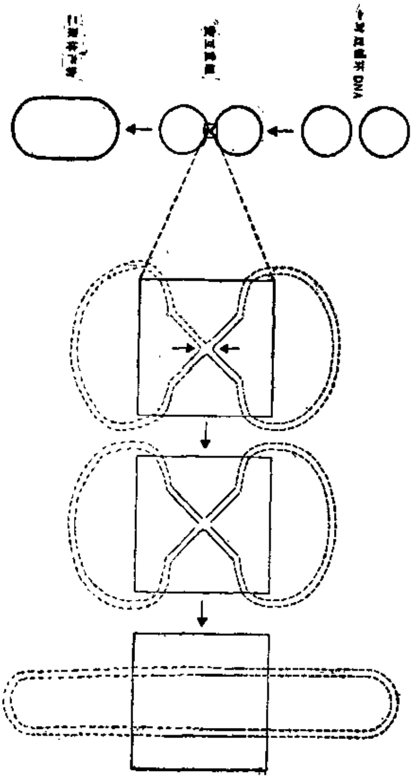


图 11-5　环形双链 DNA 分子的重组及拆分

由此可见，Holliday 中间体可以以两种交替方式进行拆分，两侧基因或是发生交互重组，或是无重组发生。但不论哪种拆分方式都必然在两条 DNA 分子上留下一段异源双链区。

上述讨论所涉及的是两条线状 DNA 双螺旋之间的重组，其原理同样适用于环状 DNA的重组过程。由于许多环状 DNA，如噬菌体和质粒等，其分子小，便于操作，而且其分子生物学和遗传学方面的研究较为透彻，因此是研究遗传重组的理想材料。图 11-5 为环状 DNA 分子之间的重组过程。图中的放大部分展示了图 11-4 相同的交叉点结构；只是这里每条

DNA 分子的首尾相连，这种结构叫做 8 字型中间物(figure-8 intermediate)。8 字型结构的拆分也有两种方式：一种是产生两个亲体环状 DNA 分子，但每个各含一段异源双链区；另一种是以交互重组方式产生一个由两个亲体 DNA 分子首尾共价连接而成的二聚环(dimeric circle)(图 11-5)。这意味着对于环状 DNA 分子，相互重组必然导致二聚环的产生；二聚环的进一步相互重组将产生多聚环。由二聚环重新产生单体环状 DNA 分子则要经过第二轮重组过程，它可以在二聚环中两个单体分子的任何同源区域之间发生，形成两个重组环状 DNA 分子(图 11-6)。

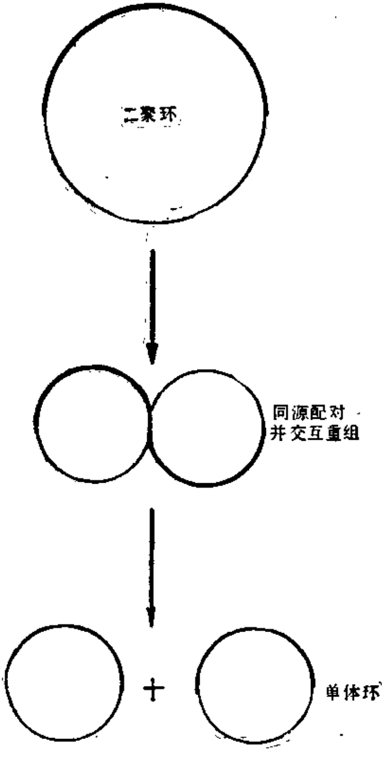



图 11-6　二聚环 DNA 分子经过相互重组而形成两个单体 DNA 分子，相互重组可以在二聚环上任意两个同源区域之间亲行。

如果图 11-5 的模型是对的，在实验中应该有可能分离到作为重组中间物的 8 字型结构。外表为 8 字型的二聚体 DNA 分子已经在好几种环状 DNA 噬菌体分别侵染的大肠杆菌细胞中发现。可是，在电镜下具有 8 字外形的结构并不一定意味着它们是以图 11-5放大部分所示的形式存在的重组中间物，其他情况如两个相互环连的(Catenated) 单体 DNA 分子在电镜下也呈 8 字形。不过，重组中间体的两个单体 DNA分子以共价连接，而环连的单体 DNA 分子之间无此连接。因此，若以一种在单体分子上有一个切点的内切酶处理，环连所涉及的两个单体将彼此分开，而重组中间体的两个单体将在交叉处仍然相连。经限制酶切割的 8 字型重组中间体具有四条臂，整个形状尤如希腊字母 $\chi$ (读作 Chi)，因此叫做 Chi 结构(Chi structure)。图 11-7 为一种环状 DNA 噬菌体的 8字形重组中间体经切割后形成的Chi 结构的电镜实际观察图解及其解释图，该图与图11-3、11-4 和 11-5 中预期出现的模式十分吻合。Chi 结构的分离成功直接验证了细胞内重组过程中异源双链 DNA 以及 Holliday 中间体及其异构化的存在。

大肠杆菌中，Chi 结构只出现在 $recA^+$ 细胞中，而不出现在 $recA^-$ 突变体中。因此，它的存在与细胞进行遗传重组的能力有直接的平行关系。这进一步说明 Chi 结构的确是遗传重组的中间物而不是与之无关的分子怪物。

环状 DNA 重组的研究进展为分离提取催化重组的酶提供了鉴定方法。测定大肠杆菌细胞提取物由单链环状分子产生二聚环的能力，发现野生型细胞($recA^+$)具备此能力。产生出可被切割成 Chi 结构的 8 字形中间物；$recA^-$ 细胞提取物则无此能力。从理论上讲可以将离体系统分级分离到各个组分，然后采用离体互补测定的方法，鉴定它们在重组中的功能。这

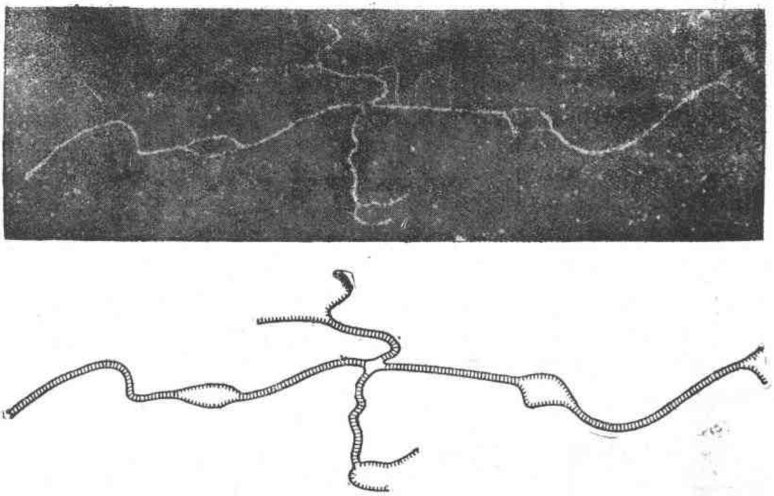


图 11-7 8字形重组中间体经限制酶切割而形成的 Chi 结构的电镜照片及解释图

类工作正在进行中。

近年来，重组中间体已经在真核生物中发现，在腺病毒侵染过的细胞中，已经在电镜下发现了一些DNA分子的一条单链正在侵入另一个DNA双螺旋分子；也发现了一些具有四条臂的 Holliday 中间体。酵母的 2 $\mu$m 环状质粒本身含有两个头-头相连的倒转重复单位。2 $\mu$m 环可以在分子内或分子间发生重组，将重组中间体用限制酶处理后，产生如同噬菌体 DNA 一样的 Chi 结构。有趣的是，Chi 结构只是出现在处于减数分裂的酵母细胞中，大约每 1000 个 2$\mu$m DNA 单体中可以观察到一个 Chi 结构（酵母每细胞约有 50～100 个2 $\mu$m 质粒）；有丝分裂细胞中则无此结构。此外，在非洲爪蟾卵母细胞的提取物中也发现有分支 DNA 分子的存在，它们很可能是 Chi 结构。

### 三、异源双链与基因转换

遗传重组中异源双链的形成及 Chi 结构的拆分解释了许多遗传学现象。处于异源双链区两侧的基因在形成重组 DNA 分子的拆分中可以发生交互重组，也可以不发生交互重组；对于处于异源双链区内的基因或遗传标记，不论拆分以何种方式发生都将影响到它们的遗传学行为。

许多子囊菌都是研究异源双链区遗传学变化的好材料。一个杂合体通过减数分裂而产生的四个核位于同一个大的细胞即子囊之中，而且这四个单倍体的核呈线性排列。紧接减数分裂之后便是一次有丝分裂，因而产生了 8 个线性排裂的子囊孢子。在不发生基因重组的情况下，具有两种不同颜色的孢子按 4:4 的比例出现(图 (11-8a)。然而，有时并不呈现预期的 4:4 分离比例，而是呈现 5:3 或 6:2 分离比例。这种现象是基因转换 (gene conversion) 而产生的。

在异源双链区存在着不配对碱基。细胞内错配修复系统(见第四章)能够识别不配对碱基，并以其中一条链为模板修复另一条链。这种修复作用发生在减数分裂和有丝分裂之间。如果在此期间，两个异源双链没有被修复，则会出现改变了孢子排列顺序的 4:4 分离比例(图 11-8b)。这个现象叫做减数分裂后分离(postmeiotic segregation)，因为它取决于减数分裂所产生的4个单倍体 DNA 双链在随后的复制中发生分离。如果只有一个异源双链得到修复，8 个子囊孢子则按 5:3 (或 3:5)的比例发生分离(图 (11-8c)。如果两个异源双链都按同一方向进行矫正，则出现 6:2 (或 2:6)的分离比例(图 11-8d)。这种通过对异源双链区内不配对碱基的修复而进行的基因矫正过程称为基因转换。

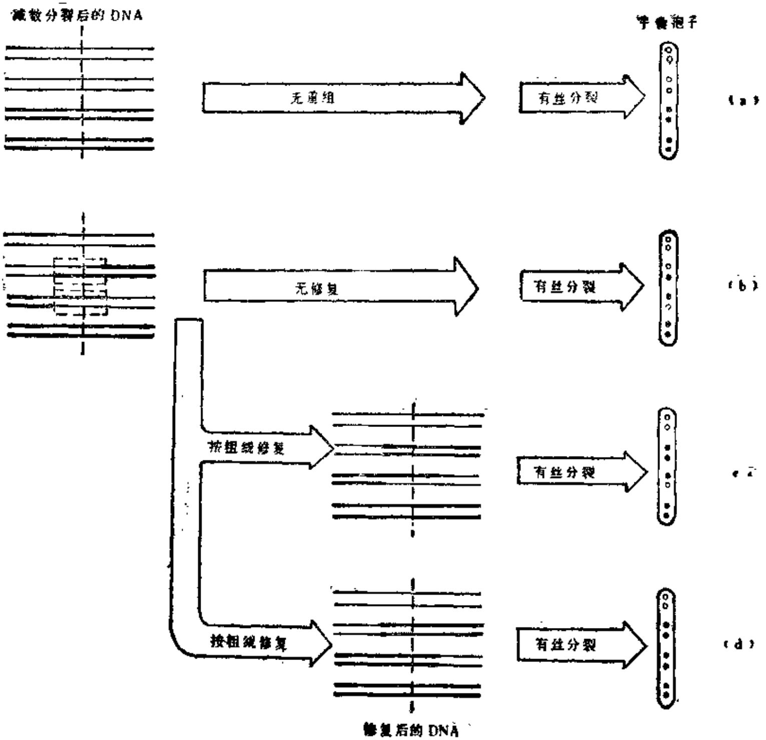



图 1-8　子囊菌孢子形成过程中的基因转换现象

(a) 无重组；(b) 有重组但无修复，在有丝分裂后分离；(c) 一个异源双链被修复，出现 5:3 的分离比例；(d) 两个异源双链按相同方向进行修复，出现 6[illegible]离比例。

基因转换并不依赖于交互重组的发生。大量的试验结果表明，子囊菌中大约一半的发生了基因转换的子囊同时伴随着两侧基因的重组。这一结果就说明了图 11-4 中重组中间体的两种拆分方式机会均等。

对于含有两点差异的等位基因，这两点之间可能发生交互重组，但因为它们处在一个基因

上，靠得很近，也可能被同一段异源双链区复盖，以基因转换的方式发生重组。所以，表现型上观察到的基因内重组并不必然伴随着两侧基因的交互重组。等位基因存在差异的两点之间的距离与基因转换中修复校正作用有一定的关系：两点相距越近，它们的共转换(coconversion)概率，即在异源双链的同一段上按相同方向同时被修复校正的概率就越大；两点相距越远，共转换概率就越小。这一特点已被用来进行基因内细微结构的制图。

回顾本节第一部分所述的 Meselson-Radding 模型，可知在重组事件的起始阶段只有一条 DNA 分子中含有非对称异源双链区，只是到后来由于异构化和分支迁移才使两条 DNA 分子上出现对称异源双链区。这一点已由近来发现的异源双链区上、下游部分基因转换频率和比例的不同而得到证实。子囊菌中的基因转换呈现一定的极性(polarity)，据此而提出的极化元(polaron)假说认为，造成基因转换的重组事件优先在某一基因一侧的特定部位起始，所形成的异源双链进入该基因，在其中的不同部位随机终止。这样，越是靠近优先起始点的遗传标记，其发生基因转换的频率就越高；越是远离该点的标记，其转换频率就越低。这样，就形成一种自起始点由高向低变化的梯度。近来在粪盘菌(Ascobolus)的 *b*2 基因中发现基因转换频率在非对称异源双链区中最高，在对称异源双链区中最低，由前者到后者呈一种递降梯度。

酵母中的基因转换也呈极性现象，但与粪盘菌等不同的是酵母的异源双链区主要是非对称性的，对称性部位只占一小部分。这反映了酵母在减数分裂形成重组中间体过程中倾向于在一条 DNA 分子上形成较长的非对称异源双链区，而在 DNA 分子异构化后，分支迁移步骤进行不久就终止了。

基因转换不仅仅在减数分裂时非姊妹染色子体的等位基因之间发生，还在以下情况下发生：(1) 有丝分裂时姊妹染色子体等位基因之间；(2) 有丝分裂和减数分裂时姊妹染色子体的非等位重复基因之间；(3) 有丝分裂和减数分裂时同一条染色子体上非等位重复基因之间。在(2)和(3)中，基因转换的频率远远高于相应的交互重组频率，正如在第二章中所阐述的，基因转换使同一家族的串联重复基因拷贝处于不断地相互“比较”之中，在维持这些重复单位的同一性方面起了重要作用。

### 四、细菌转化、接合和转导的重组机制

以上讨论的是两个完整的双螺旋 DNA 分子(线状的与环状的)之间的重组。重组也可以发生在一个完整双螺旋 DNA 分子和一个 DNA 片段(单链的或双链的)之间。例子有细菌的转化，接合和转导等。

1. 细菌转化　细菌转化中供体 DNA 片段进入受体细胞和受体染色体发生重组。肺炎球菌等许多细菌的转化DNA只有一条链进入受体细胞，另一条链被分解掉。因此，实际上重组发生在单链DNA 片段和完整的双链 DNA 之间。在受体细胞内，刚刚进入的供体单链 DNA 和 SSB 蛋白质结合而免遭核酸酶的攻击，然后，供体 DNA 取代受体染色体上的一段同源序列，形成局部异源双链区(图 11-9a)。目前流行的解释这一过程的模型见图 11-9b。在类似大肠杆菌 RecA 蛋白质的某种因子协助下，供体单链 DNA 侵入受体 DNA 上与其同源的区段，取代同源单链并与互补链形成碱基配对。DNA 呼吸现象所提供的暂时性解链可能为这一侵入过程提供了方便，通过尚不清楚的机制被取代的受体单链遭到切割，继而出现分支迁移使供体单链 DNA 取代区域逐步扩大；同时外切酶系统不断切去供体单链上和受体被取代单链上的游离端。结合进去的供体单链和受体 DNA 之间的切口由 DNA 连接酶连接，形成一段异源

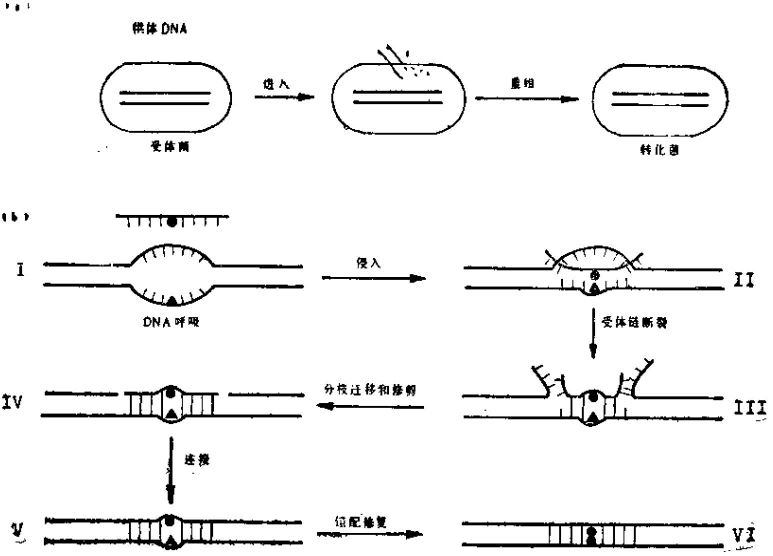


图 11-9 细菌转化的可能机制
(a) 转化体产生的大致过程； (b) 转化中遗传重组的可能机制

双链区。如果所涉及的 DNA 序列在供体和受体中不完全一致，错配碱基对便由此产生。

转化过程的最后结局取决于错配碱基对的修复校正。如果校正时被切除的是异源双链区中原属供体单链的碱基，则实际上无重组发生。如果被切除的是原属受体 DNA的碱基，则重组发生，该细菌的后代细胞全部表现为重组体。如果无修复校正作用发生，则该细胞经 DNA 复制和细胞分裂后产生的两个细胞中，一个具有受体的基因型，另一个具有重组体基因型。但是由于转化试验中所采用的选择条件一般只允许重组体细胞生长，故在无校正作用时产生的菌落中绝大多数细胞为重组体。有些遗传标记在转化中或是很少发生校正作用，或是校正切除几乎总是在受体DNA 上。因此转化频率较高，这类遗传标记中叫做高效率标记(high-efficiency markers)。而另一些遗传标记在转化中的校正切除总是倾向于发生在供体单链上，因而表现出很低的转化频率。这类标记叫做低效率标记(low-efficiency markers)。可以推测，一旦细胞内的碱基错配校正功能失活，低效率标记将转变为高效率标记。近来已分离出碱基错配修复缺陷型突变体，以它们为受体进行转化实验时，所有供体遗传标记的转化效率都非常高。这一结果论证了上述解释的正确性。

2. 细菌接合　在以Hfr 细胞为供体、$F^-$细胞为受体的大肠杆菌接合过程中，供体DNA以滚环方式产生单链 DNA 进入受体细胞并重组到受体 DNA 中，接合时 DNA 重组的机制类似于转化，但过程更为复杂。供体DNA 单链一般比转化 DNA 的单链长得多，若全部进入受体细胞，原则上讲可以同时和整个受体 DNA 配对并发生重组，当然实际上往往只有一部分供体DNA 能进入受体，而进入的部分中也只有一部分能被结合到异源双链中去。大肠杆菌的接合需要 RecA 蛋白质和 *recB*、*recC* 及 *recD* 基因的产物 RecBCD 蛋白质。

3. 转导　和转化与接合不同，无论是普遍性转导还是特异转导的遗传重组都在双链DNA片段和完整DNA分子之间发生。转导DNA以双链的形式被注入受体细菌，然后同样以双链的形式结合进受体染色体中，这一过程很可能如同两个完整DNA分子的重组那样首先起始于单链侵入和取代，然后两条链都被牵扯进去。转导也需要RecA和RecBCD蛋白质。目前对转导中的重组机制了解得甚少。

## 第三节　同源重组的酶学机制——RecA蛋白质和RecBCD蛋白质在同源重组中的作用

### 一、RecA蛋白质在联会和链交换中的作用

大肠杆菌的RecA蛋白质是一个39 KDa的单肽链，是同源重组过程中最重要的蛋白质，因此有人称它为重组酶(recombinase)。

近年来，对大肠杆菌RecA蛋白质在离体条件下促进类似于体内重组过程的研究正在使人们对同源遗传重组机制的认识不断深化而进入分子水平。RecA蛋白质在遗传重组中的作用主要包括促进同源DNA联会和促进DNA分子间的单链交换。

RecA蛋白质是SOS反应中的关键调节蛋白之一，这已在第四章第二节中提到。RecA蛋白质的这一特点主要是基于它在DNA损伤或复制受阻的情况下促进特异性的蛋白质的分解。RecA蛋白质还有许多与重组有关的活性，其中主要包括。

1. 单链DNA结合活性　在中性pH的条件下，RecA能够大量地结合于单链DNA，而且这种结合呈现高度的协同效应。在有足够的RecA时，能将单链DNA全部占据。平均一个RecA单体能与4个核苷酸结合(各种文献中有的低至3个，有的高至6个)。RecA与单链DNA结合后，这个复合物的半径大约为12 nm，其轮廓长度大约为自由双链DNA的长度的60%。但是在ATP存在时，RecA-单链DNA复合物的长度可以长达自由双链DNA长度的150%。这样的复合物很可能就是RecA所促进的DNA单链交换过程中的活性形式。

RecA蛋白质与单链DNA结合是相当稳定的，其半寿期大约为30分钟，但这种稳定性受阴离子的种类和核苷酸辅因子的强烈影响。当ATP加入到RecA-单链DNA复合物中时，其半寿期缩短至3分钟，当ADP加入时，其半寿期缩短得更加厉害，大约为0.2分钟。然而ADP的存在并不影响RecA蛋白质与单链DNA的结合倾向，只不过是加快结合与解离的平衡。与ATP相反，ATPγS的加入并不能使RecA从它与单链DNA的复合物中解离。这就表明，在体内，当ATP存在时，促进了RecA与单链的结合，然后由于RecA具有依赖于单链DNA的ATPase活性，分解了ATP而释放出ADP，又由ADP促进RecA脱离已经发生链交换的单链DNA，再由ATP的促进转移到其他单链DNA上。

2. 双链DNA结合活性　RecA蛋白质与双链DNA的结合不同于它与单链DNA的结合：(1) 高度的pH依赖性，最适pH为6.0，在pH 7.5时结合活性已难于观察到，(2) 这种结合只有在ATP等核苷三磷酸存在时才发生。由于RecA的结合，双螺旋DNA发生解链，这样形成一种较粗的棒状螺旋形复合物，其直径大约100Å，螺距亦为100Å左右，其中包含18.6碱基对，结合有6.3个RecA分子。RecA与双螺旋DNA的结合以及DNA的解链，并不伴随着ATP的大规模水解，而是一个相当缓慢的过程，最终导致了RecA与单链DNA

的结合状态。以便进行链交换，在体内，很可能 RecA 蛋白质与双螺旋 DNA 结合，使 DNA 接近于这种转变状态。

3. NTP 酶活性 在单链或双链 DNA 存在时，RecA 蛋白质能够水解 ATP(dATP)，GTP(dGTP)，UTP(dUTP)，水解 CTP(dCTP) 的活性较低。水解 TTP 的活性更低。与单链或双链 DNA 结合活性相一致的是：单链 DNA 存在时，可在较宽的 pH范围内发挥NTPase 活性，最适 pH 为 8.0；而在双链 DNA 存在时，仅在极狭的 pH 范围内 (pH 5.5—6.5) 具有 NTPase 活性，最适 pH 为 6.2。然而，只有 ATP 的水解才能促进联会。

用 UTP 抑制试验表明，RecA 分子只有一个核苷三磷酸结合位点。而与此相反，RecA分子可能有两个 DNA 结合位点(见下述)。

4. 促进互补单链复性的能力 在 ATPγS 存在时，一个 RecA 分子能够结合 8 个单链 DNA 的核苷酸。这就意味着，RecA 分子可能具有两个核苷酸结合位点。这样，两条单链可能通过 RecA 分子相互连接(DNA·RecA·DNA)。这样的模式可用下述试验证明：当 RecA 蛋白与 $\phi$X174 正链 DNA(线形)结合以后再加负链 DNA(线形)，另一种是将 $\phi$X174 的正链(线形)和负链(线形)以及RecA蛋白质一起混合。这样，两种情况下形成RFII形式所需的时间大体相同。当然，还不能排除另一种可能性即通过形成 DNA·RecA·RecA·DNA 而促进单链 DNA 的复性，因为 RecA 分子本身也有相互结合成丝状复合物的能力。在复性过程中，碱基互补当然是不可缺少的条件。RecA 所促进的单链 DNA 复性与 SSB 所促进的复性完全不同：

(1) RecA 所促进的单链 DNA 复性是在 RecA 亚饱和状态下进行的，大约 1 分子 RecA 比 30 个单链 DNA 核苷酸，饱和状态的 RecA 会抑制复性。SSB 却与此相反。

(2) SSB 促进的单链 DNA 复性，可能是防止链内二级结构形成，从而增加了"成核"的机会，然后通过拉拉链作用实现复性，分子碰撞仍然是其限制步骤，因此为二级反应(见第一章)；而 recA 促进的单链 DNA 复性为一级反应，表明在复性之前有一个迅速的非限制性的中间复合物形成过程。

(3) RecA 蛋白质所催化的复性强烈地受到 ATP 的刺激，而 SSB促进的复性不需要任何核苷三磷酸。

上述关于 RecA 蛋白质的四种性质，在体内则汇聚于一个更为复杂和更为完整的反应之中，那就是促进 DNA 分子的同源联会和 DNA 分子之间的单链交换。

RecA 蛋白质促进单链 DNA 和双螺旋 DNA 之间的同源联会，参与联会的可以是单链 DNA 片断，双链 DNA 片断，线状的完整单(双)链 DNA、环状的单(双)链 DNA；双链 DNA 上可以带有也可以不带有单链切口或单链末端(图11-10)。但无论哪种组合，其中的一方必须是单链分子或带有部分单链区。同源联会的一般途径如图 11-11 所示，全过程可以分为三个阶段，即联会前、联会和链交换。

1. 联会前阶段 首先 RecA 蛋白质与单链 DNA 区域结合形成丝状复合物。RecA蛋白质的需要量取决于单链 DNA 的长度，二者呈化学计量关系。当每 3—5 个核苷酸的单链 DNA 对应一个 RecA 蛋白质分子时，同源联会效率最高。

2. 联会阶段 RecA-单链 DNA 复合物继而与双链 DNA 结合，形成 RecA-单链 DNA-双链 DNA 三元复合物。这种结合与单、双链 DNA 序列的同源性无关，仅靠开始形成复合物时的随机碰撞。单链 DNA 和双链 DNA 之间很难做到同源序列的正好对齐，还需进一步反应

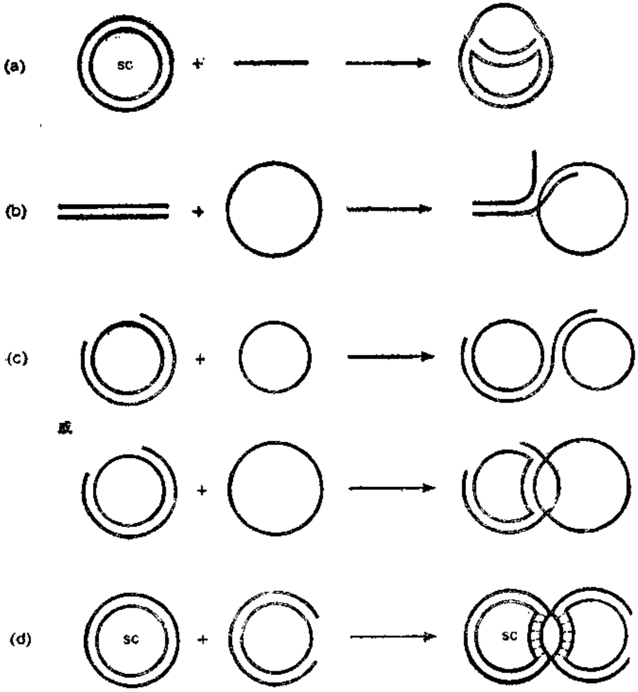



图 11-10 RecA 蛋白质促进的各类 DNA 分子之间的联会

才能做到这一点。一种可能性是 RecA-单链 DNA 与双链 DNA 之间极其迅速地反复形成三元复合物，然后随即破坏，直至找到同源序列；另一种可能是在三元复合物内单链 DNA 和双链 DNA 之间以某种方式相错运动，直至“寻找到”同源序列。目前尚不知双链 DNA 是如何解链以允许同源性的测试的。

3 链交换阶段 当单链 DNA 和双链 DNA 在 RecA 蛋白质的促进下实现同源配对后，单链 DNA 便能侵入双链 DNA 将其中的同源单链置换出来，自己与互补链进行碱基配对，这一反应叫做单链摄入 (single strand uptake) 或单链同化，它与本章第二节图 11-2 中所述的步骤相似。单链同化导致了D-嗜噗的产生(图 11-11)。这一过程需要 ATP 的水解。由 RecA 蛋白质行使 ATP酶活性，水解一分子的ATP可以促进5—10个核苷酸的同化。单链同化所形成的异源双链区随同化作用而逐步扩大，该效应类似于前面所述

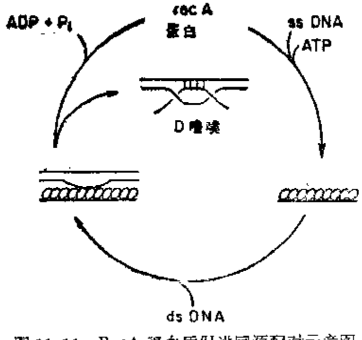



图 11-11 RecA 蛋白质促进同源配对示意图

的分枝迁移。但在无RecA 蛋白质的情况下，D-噜噗的分枝迁移是双方向性的，由DNA呼吸作用控制，位移速度极快，可达每秒 80,000bp。而由RecA蛋白质催化的单链同化是单方向性的，速度极慢，每秒钟仅几个 bp。不过单链同化可以允许1% 以上的错配碱基对出现，而 D-噜噗的分枝迁移难以越过一个错配碱基对。

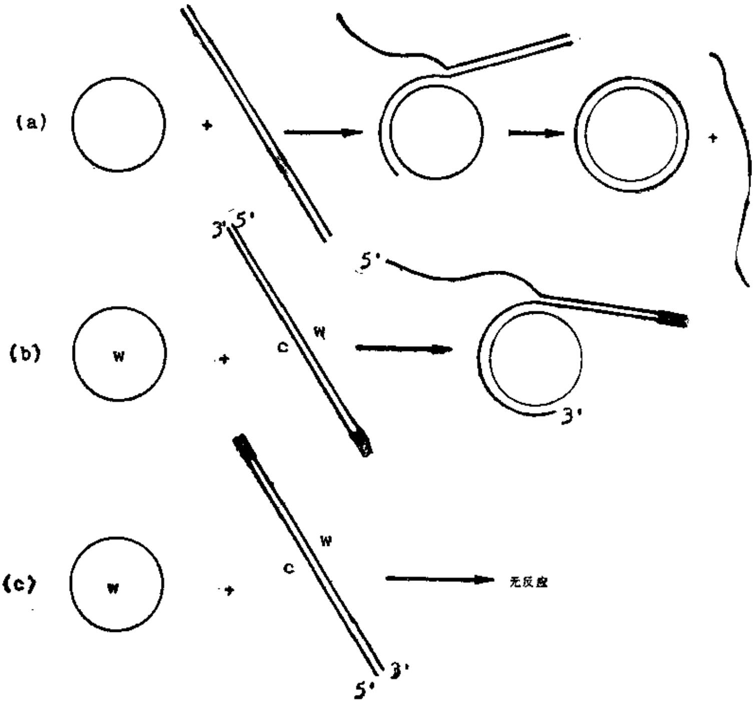



图 11-12 (a) RecA 蛋白质促进双链线状 DNA 与单链环状 DNA 的链交换；(b)和(c)说明了这一反应中的方向性，二条等长的粗线表示不同源的异种 DNA 序列；W 和 C 为互补的单链 DNA。

链交换的单方向性可以由以下实验得到证实。将闭合环状单链 DNA 和同源的线状双链 DNA 在 RecA 蛋白质催化下重组。双链 DNA 中与环状单链 DNA 互补的链能够从一端开始与单链环状 DNA 配对，形成在性质上和 D-噜噗等同的连锁分子(joint molecule)。然后，通过单链同化作用，异源双链区越来越长，直至把原来双链 DNA 中和单链环状 DNA 同源的单链整个置换出来，其长度可达数千 bp(图 11-12 a)。人们发现，若在线状双链 DNA 的不同一端共价连接一段与单链环状 DNA 完全不同源的异种双链 DNA (heterologous DNA)，对链交换的效应截然不同；如果异种 DNA 连接在双链 DNA 上与单链 DNA互补的那条链的 5′端一侧(图 11-12 b)则链交换照常发生，异源双链一直形成到异种 DNA 处才停止；如果异种 DNA 连接的一端相当于单链 DNA 互补链之 3′ 端，则无链交换出现(图 11-12 c)。这说明单

链同化只能沿固定方向进行，这个方向对双链 DNA 上的互补链来说是 3′→5′；对入侵的单链(在这里是单链环状 DNA)来说是 5′→3′。其他实验结果表明，连锁分子形成起始阶段无方向性，而单链同化则有方向性。

链交换过程中异源双链的形成与双方序列的同源性有密切关系。φX174DNA 和 G4DNA 有 70% 的同源性，在 RecA 蛋白质存在时，φX174 DNA 的单链片段不能与非超螺旋的 G4 DNA 双螺旋之间形成 D-噜噗；不过如果 G4 DNA 双螺旋呈负超螺旋状态，则可以形成 D-噜噗，产生数百 bp 长的异源双链区。其中平均约 30% 为错配碱基对。而如果不加 RecA 蛋白质，则即使在 G4 DNA 为负超螺旋的情况下也无异源双链形成。这一结果表明 RecA 蛋白质的存在可以容忍异源双链中较多的错配碱基，同时也表明双链DNA 的负超螺旋状态对异源双链的形成有促进作用。目前对这些效应的解释是：RecA 蛋白质作为解旋酶为 φX174 和 G4 DNA 的链交换提供了方便；但所形成的异源双链区极不稳定，需要负超螺旋通过内部贮存能量的释放使之稳定。

如果以序列同源性高达 97% 的 fd DNA 和 M13DNA 进行重组反应，RecA 蛋白质可以在双链处在非超螺旋的状态下催化连锁分子的形成。所形成的异源双链区包括许多错配碱基对，不过，fd DNA 和 M13DNA的链交换不能进行彻底；异源双链区的伸长可以顺利通过所有的单一出现的错配碱基对，但不能通过连续两个或两个以上的错配碱基对。

RecA 蛋白质促进含有众多错配碱基对的异源双链区形成的能力表明，两段同源性不是很精确的 DNA 也可以发生重组，重组可能伴随富含错配碱基对的异源双链区的产生，由此造成大量的基因转换事件，在 RecA 蛋白质及单链 DNA 片段的大量存在下，染色体的许多区域可能出现非同源序列之间的链交换，这可能是 SOS 反应中诱变作用的原因之一。

这里提一下 SSB 蛋白质的作用，它们单独存在时不能促进 D-噜噗的形成。但可以帮助 RecA 蛋白质促进形成 D-噗噜。当 SSB 蛋白质以饱和量(即足以和所有单链 DNA 结合)存在时，RecA 蛋白质的需要量只是原来的 1%，而且对 ATP 的消耗也大大减少。但过量 SSB 蛋白质对 D-噗噜的形成有抑制作用。SSB 蛋白质对 RecA 蛋白质的功能的促进作用可能是基于这两种蛋白质的相互作用，其机制目前尚不清楚。

DNA 旋转酶和 DNA 拓扑异构酶 I 对 RecA 蛋白质催化的链交换作用也十分重要。离体条件下，DNA 旋转酶的抑制剂能够使重组频率降低 70—80%。这些酶大概与促进异源双链的形成和解决重组时出现的拓扑学问题有关。

### 二、RecA 蛋白质和 Holliday 中间体的形成

真核生物细胞内的遗传重组发生在两条线状 DNA 之间，细菌转导和某些噬菌体的重组等也发生在两条双链 DNA 之间，那么 RecA 蛋白质离体催化同源重组的功能能否模拟这一情况，形成如第二节所述的 Holliday 中间物呢？实验证明，两条同源的线状双链 DNA 分子中只要一条含有一段单链区，就能导致 Holliday 中间体的产生。

当单链区域出现在一条线状双链 DNA 分子的末端时，在 RecA 的催化下，单链区域取代另一条分子上的同源链而与互补链进行碱基配对，形成局部异源双链区，使原来的双链 DNA 末端呈被置换的单链状态。随着链同化作用的发展，异源双链区进一步扩大，这样就为另一对互补单链之间的配对提供了可能，于是便出现对称的第二条异源双链区，然后转化成Holliday 中间体结构(图 11-13)。如第二节所述，Holliday中间体能够经过分子内异构变化而形成

Chi 结构。这种结构已通过电镜在 RecA 蛋白质催化的线状双链 DNA之间的离体重组中发现。在图 11-13b 中，过去由于画法的不同，即将双链中与单链互补的一段拉过去与单链部分形成配对，这样就形成所谓 “nu” 结构。

关于RecA蛋白质如何促进 Holliday 中间体中的分枝迁移的问题，现在仍然不清楚。可能RecA 分子结合于一条双螺旋 DNA，使其处于松缠状态，在其大沟中暴露可以进行链交换的活性部位。链交换的实质就是原有双螺旋中两条链之间的氢键断裂，再分别与另一条双螺旋中的互补链形成氢键。图 11-14a 中表示没有结合 RecA 蛋白质的一条双螺旋绕着有 RecA 蛋白质的一条双螺旋旋转，产生链交换和分枝迁移。图 11-14b表示在 Holliday 中间体中分枝点的一侧的两个双螺旋 DNA 分子形成一段四股螺旋区域，在这一区域发生了链交换活动。这两个模型都还缺乏过硬的实验证据，只是从推理上来说是行得通的。但是不管哪一种模型，都需要 DNA 拓扑异构酶来解决链交换和分枝迁移过程中所产生的拓扑学问题。实验表明，在离体条件下，DNA 旋转酶的抑制剂能使 DNA 重组能力降低 70—80%。

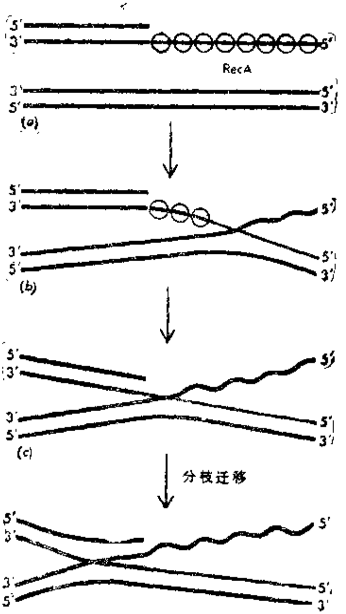



图 11-13 在体外试验中RecA蛋白质促进的两条线状双链DNA 之间的联会和链交换。其中的一条 DNA 在末端有单链区

过去所提出的有关遗传重组机制的模型(包括第二节中介绍的链交换模型)主要是基于遗传学尤其是减数分裂时基因转换的分析。对大肠杆菌 RecA 蛋白质作用的研究为了解同源重组的一般机制提供了新的认识。首先，所有生物中的同源重组可能都需要 RecA 蛋白质或类似的蛋白质催化。除了已经在许多细菌中发现了 RecA 蛋白质外，现在还在一些真核生物中发现了功能上与大肠杆菌的RecA十分相似的蛋白质。使它们失活的突变造成遗传重组的缺陷型。其次，单链DNA在 RecA 蛋白质离体催化同源重组中的突出作用验证了它在细胞内遗传重组中的重要性。此外，RecA 蛋白质在单链同化中具有单方向性，这一点可以部分解释基因转换中的极性现象。

### 三、RecBCD 在同源重组中的作用和重组热点

RecA 蛋白质无疑是遗传重组中最重要的蛋白质之一，可是它只能催化同源联会和链交换而不能催化遗传重组过程中的其它步骤，包括单链 DNA 区域的形成，连锁分子(Holliday 中间体等)的拆分。细胞内遗传重组的全过程需要其它基因产物的存在。根据对大肠杆菌各种重组突变体的研究，发现同源重组的实现除了这个最重要的称做重组酶 (recombinase)的 RecA 蛋白之外，还需要基因 *recB*，*recC*，*recD*，*recE*，*recF*，*recG*，*recJ*，*recK*，*recL* 和 *recN* 的产物。目前，只有 RecA，RecB，RecC，RecD 已经分离纯化出来。除了 RecA 之外，其余三种基因产物在同源重组中的作用也是很重要的。

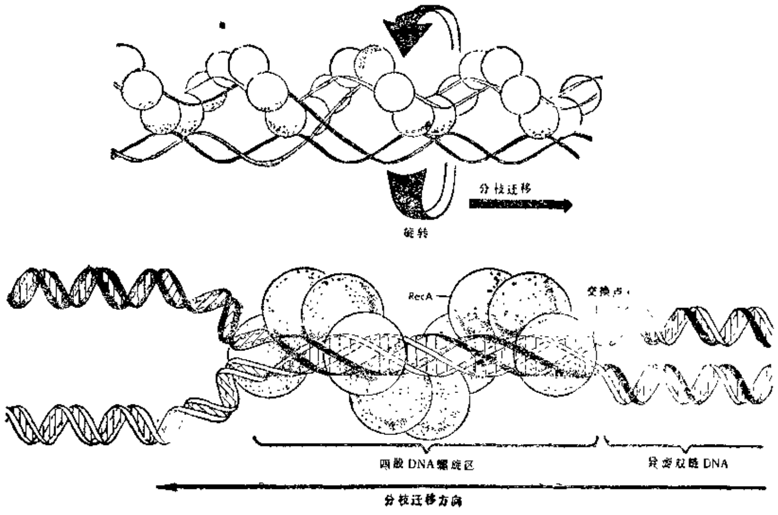



图 11-14 RecA 蛋白质促进链交换和分枝迁移的作用模型
(a) 一条双螺旋绕着另一条双螺旋旋转模型；(b) 四股螺旋模型

大肠杆菌的 *recB*, *recC* 和 *recD* 基因分别编码 130KDa, 120KDa 和 60KDa 的多肽链，三者构成一个在同源重组中的功能单位——RecBCD 蛋白质。RecBCD 蛋白质是一个多功能的酶，它具有依赖于 ATP 的单链和双链外切酶的性质，因此又称为外切核酸酶 V。它利用水解 ATP 所释放的能量，又具有线形 DNA 的螺旋酶活性。此外，RecBCD 蛋白质还是序列特异性的单链内切酶（见下述）。对于 RecBCD 酶中的三种亚基的功能最近已有初步了解。其中亚基 RecD 只到 1986 年才被发现（在此之前，人们称 RecBCD 蛋白质为 RecBC 蛋白质）。根据突变体的研究，RecB 具有依赖于 DNA 的 ATP 酶活性并具有螺旋酶活性（RecC 可能会加强这种活性）；RecD 具有核酸酶活性，包括外切核酸酶和特异的单链内切核酸酶活性。各种亚基功能的细节还有待于各种体外重建试验的研究。

RecBCD 酶在同源重组中到底有什么作用？这必须从它的螺旋酶活性和核酸酶活性入手。在离体试验中，如果缺少 SSB，RecBCD 能从线形双链 DNA 末端起始进攻，解旋 1000 bp 左右，其中一股被切成 4～5 碱基的寡核苷酸，另一股成为 1000 个碱基左右的单链尾巴。但是在 SSB 存在时，单链片断的产生极大地减少，外切核酸酶活性（包括线形单链 DNA 的外切活性）也受到抑制。在较高的 ATP 浓度（3～5mmol/L）时更是如此。这更符合体内情况（存在着大量的SSB 和较高浓度的ATP）。RecBCD对线形双链DNA的外切活性最高；也能作用于含有5个核苷酸以上单链缺口的双链 DNA，但活性比线形双链 DNA 作底物时低 10 倍。这很可能是首先由内切核酸酶切断单链后才作为 RecBCD 外切酶的底物。RecBCD 酶对于平头末端的双链 DNA 的螺旋酶活性最高，当单链末端长达 25 核苷酸以上时，RecBCD 几乎不能使之解旋。除了解旋酶活性外，RecBCD 还有再旋酶活性，但比解旋酶活性弱得多。

这样，当细胞内 DNA 具有平头末端时（这个平头末端是如何产生的？至今还不清楚），RecBCD 就结合上去。它利用 ATPase 活性水解 ATP，从 DNA 一端将双链 DNA 解开，并向另一端移动。由于再旋酶活性较低，而且解旋位点和再旋位点由同一个酶复合体所固定，因此形成所谓兔耳结构，也就是两个单链噜噗结构（图 11-15）。当 RecBCD 所识别的单链位点即所谓 Chi 位点（5′ GCTGGTGG 3′）进入单链的噜噗区域时，RecBCD 的特异性单链内切酶活性就在 Chi 位点的 3′ 方向 4～6 个核苷酸处将单链切断。这样，当 RecBCD 继续前进时，则留下一条包含 Chi 位点在内的单链尾巴和一段单链缺口。其结果是 RecA 蛋白质结合于这个单链尾巴，然后与同源序列进行链交换。这一过程示于图 11-15。

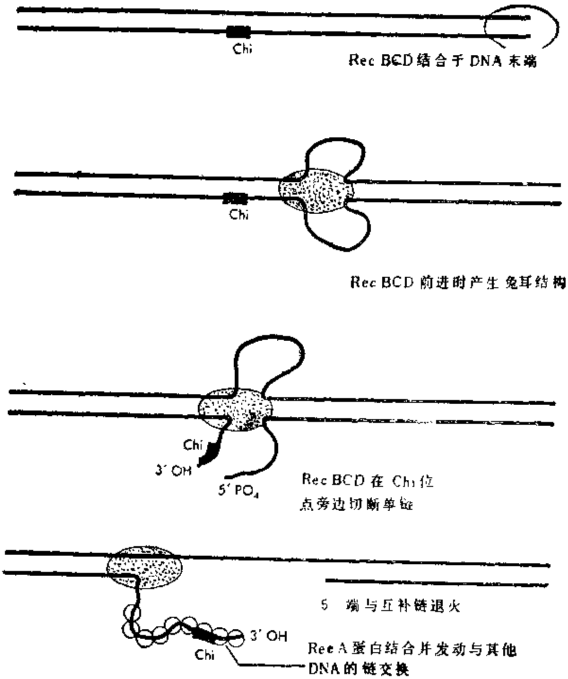



图 11-15　RecBCD 蛋白质作用双链 DNA 末端，在 Chi 位点附近产生单链 DNA，继而由 RecA 蛋白促进同源重组。

RecA 蛋白质催化的同源重组从原则上讲可以发生在 DNA 的任何同源序列之间。但是在实际上某些序列中重组发生的频率要远远高于其他序列。也就是说，DNA 分子上含有重组几率较多的热点。目前发现的重组热点主要就是 Chi 位点。Chi 位点最初是在突变的 λ 噬菌体中发现的。

λ 噬菌体在裂解生长中的 DNA 复制有两种方式，一种是 θ 式复制，在 DNA 复制的最初阶段采用；另一种是滚环式，在复制后期采用。由 θ 式复制到滚环式复制的转变需要 λ 噬菌体

基因 *red* 和 *gam* 的产物存在。*gam* 基因产物的作用之一是抑制寄主 RecBCD 蛋白质的外切酶活性；没有 *gam* 基因产物，RecBCD 蛋白质就会妨碍由 $\theta$ 式复制到滚环式复制的转变。λ噬菌体的成熟需要形成 DNA 多联体，它可以由滚环式复制方式提供，也可以由 $\theta$ 式复制所形成的单体 λDNA 经过重组形成二聚环，再逐步形成多聚环(即多联体)(参见图 11-5)。$gam^-$突变体中因不能进行滚环式复制，多联体的制造不得不全依赖 DNA 重组过程。噬菌体释放量因此大大降低。λ 的 *red* 基因产物负责 λDNA 的遗传重组。$red^-gam^-$双重突变体不但使DNA复制只局限于采用 $\theta$ 方式，而且失去了噬菌体自身编码的重组系统，这时多联体的产生和 λDNA 的正常遗传重组必须依赖寄主的 RecA 重组系统。人们发现，当 $red^-$ $gam^-$ 噬菌体双重突变体中出现另外一些突变时，该噬菌体的释放量和重组频率大大增加。对这些突变体的 DNA 进行序列分析，发现它们都是单碱基替代的点突变，使突变了的 λDNA 中出现野生型所没有的 8 bp 核苷酸序列 5′ GCTGGTGG 3′。这一序列被人们称为 Chi 序列。

大肠杆菌含有大约 1000 个 Chi 位点，亦即每五个基因就有一个Chi位点，这就给RecBCD 酶提供了许多作用位点。然而，这些位点通常是很难得到利用的，因为大肠杆菌染色体没有自由末端。在细菌接合过程中，Chi 位点可能起重要作用。自由末端可由雄性细菌产生，RecBCD 结合上去，直到发现 Chi 位点并切断，然后由 RecA 蛋白质促进与受体菌的同源重组。

近来人们发现许多低等和高等真核生物的 DNA 中都存在着 Chi 位点，而且有证据表明，这些 Chi 位点中很多能够作为重组热点而刺激遗传重组的发生。

## 四、拆分Holliday 中间体的酶活性

同源重组的最后一步就是 Holliday 中间体的拆分。人们根据 RecBCD 的核酸酶性质，推测 RecBCD 在 Holliday 中间体的拆分中起重要作用，但是目前还缺乏直接的证据。最近人们发现 T4 噬菌体基因 49 所编码的内切核酸酶 Ⅶ 能够切割 Holliday 中间体的类似物——十字形 DNA (见第一章)。这种酶能对称地在离十字形端点的 5′ 方向 2～3 核苷酸处切开十字形 DNA。在酵母的抽提液中也发现了类似的酶活性。

## 五、RecF 途径

遗传学分析表明，$recA^-$ 突变使大肠杆菌细胞内遗传重组频率降低 $10^6$ 倍；当 *recA* 基因正常时，$recBCD^-$ 突变使重组频率降低 $10^2$ 倍，这说明 RecBCD 途径是正常细胞中遗传重组的主要途径。$recA^+$ $recBCD^-$ 细胞中残存的重组活性受许多基因控制，其中较为重要的有 *recE*、*recF*、*sbcA*、*sbcB* 等。正常细胞中 *recF* 途径的重组效率非常低，但在 $recBCD^-$ $sbcB^-$ 双重突变体中这一效率可以大大提高。已知 *sbcB* 基因编码外切核酸酶 I(ExoI)，该基因突变为什么能提高 RecF 途径的重组效率，至今仍不很清楚。*recE* 编码外切核酸酶 Ⅶ，能提高 RecF 途径的重组效率，但是 *recE* 基因活性又被 *sbcA* 基因产物所阻遏。因此，大肠杆菌细胞内同源重组有两条途径：主要途径为 RecBCD 途径，次要途径为 RecF 途径。两个途径的关系和每个途径自身都是极为复杂的。虽然有人推测其中的关系并作了图解，但是其中需要证实的问题比说明的问题还要多。

## 第四节 依赖于同源重组的位点特异性的序列代换——酵母 MAT 序列的转换[19~22]

在第九章中，我们曾经谈到过，酵母接合型的转变是 DNA 序列的代换，而不是互换。其实验证据是：如果在活跃匣子的 Y 区域引入一个突变标记，而在接合型转变之后，这个突变标记则永远消失。也就是说，被代换的序列的命运是被降解掉。现在已经弄清，由 HO 基因所编码的内切酶在 *Y/Zl* 交界处有一个识别位点，其序列为 CAGCTTTCCGCAACAGTAAA。HO 内切酶在该序列内产生双链断裂，正是这种双链断裂(而不是单链切口)起始了 MAT 序列的转换过程。HO 内切酶只能作用于 MAT 匣子上的识别位点。虽然沉寂匣子上也有同样的位点，HO 内切酶却不起作用。这仍然是由于阻遏蛋白 Sir 的作用使得沉寂匣子中的这段染色质浓缩而变得致密。Y/Zl 交界处(实际上是在 Zl 区左边 10bp)的双链断裂后，Y 区域就被降解(哪种酶降解这一区域还不清楚)，一直进行到$Y_\alpha$和Ya都相同的部分或者一直到X区域。这样 MAT 部分的四个断头就先后侵入供体匣子(沉寂匣子)的同源部分并与其中的互补链配对。再以供体 DNA 为模板复制 Y 区域，形成 Holliday 中间体，然后通过拆分，产生新的 MAT 匣子和这次转换事件中的供体匣子。这个过程示于图 11-16。

酵母 MAT 匣子的代换，强烈地依赖于序列的同源性，而且是大范围的序列同源性；同时还需要重组酶系统的参加，如 *rad* 52 基因产物；另外需要位点特异性的 HO 内切酶。因此这一事件，既具有同源重组的特征，又具有位点特异性重组的特征。从遗传信息转移的角度来说，也可以算作转座过程；然而这种说法过于牵强：它既无末端倒转重复，又无靶点序列的倍生，也不形成转座的中间体。总之，酵母 MAT 匣子的转换是一个具有多方面特点的重组过程。

## 第五节 位点特异性重组[23~26]

### 一、λ 噬菌体 DNA 的整合与切除

λ 噬菌体侵入大肠杆菌细胞后，面临着裂解生长还是溶源生长的选择。进入溶源状态需要 λDNA 整合进寄主 DNA。由溶源状态进入裂解生长需要 λDNA 从寄主 DNA 上切除下来。这里，整合和切除均通过细菌 DNA 和 λDNA 上特定位点之间的重组而实现。这些特定位点叫做附着点(attachment site)。简写为 att。

大肠杆菌 DNA 上的 *att* 位点叫做 *att*B(又叫做 *attλ*)，位于 *bio* 和 *gal* 操纵元之间，含有 B、O 和 B′ 三个序列组分；λ 噬菌体 DNA 上的 *att* 位点叫做 *attP*，由 P、O 和 P′ 三个序列组分构成(图 11-17)。*attB* 和 *attP* 中的 B、B′、P、P′ 序列大不相同，而 O 序列则完全一致，是位点特异性重组发生的地方。由于线状的 λDNA 在侵入细胞后不久就已经首尾连接成环了，所以在 att 处的相互重组导致了整个 λDNA 整合进寄主 DNA(图 11-17)。在整合状态下，λ 噬菌体呈线状，两边各有一个 *att* 位点，这两个位点是重组的产物，不同于原来的 *attB* 和*attP*。原噬菌体左边的是 *attL*，由序列组分 BOP′ 组成，右边的是 *attR*，由序列组分 POB′ 组成。有时把整合反应写成：

$$\underset{\text{细菌}}{BOB'} + \underset{\text{噬菌体}}{POP'} \xrightarrow[\text{IHF}]{\text{Int}} \underset{\text{原噬菌体}}{BOP' + POB'}$$

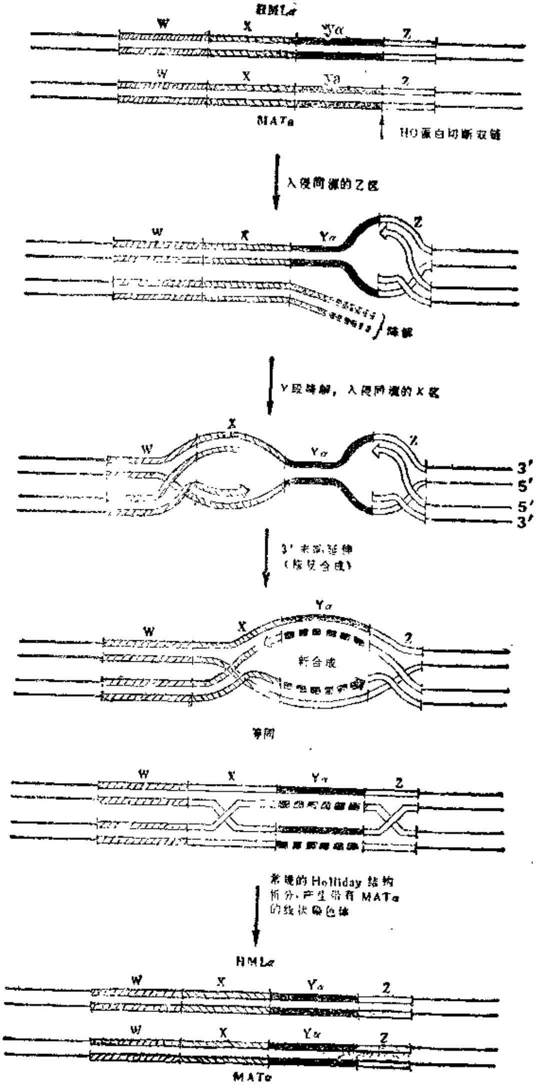



图 11-16 酵母 MAT 匣子转换的机制。首先 MAT 上产生双链断裂，然后是 MATY 片段的降解。X 片断剩 Z 片段的自由末端侵入供体匣子的同源区域。通过3′末端的延伸而形成 Holliday 中间体。最后通过拆分而完成 MAT 匣子转换的全过程。注意MAT匣子和 HM 匣子处于同一条染色体上。

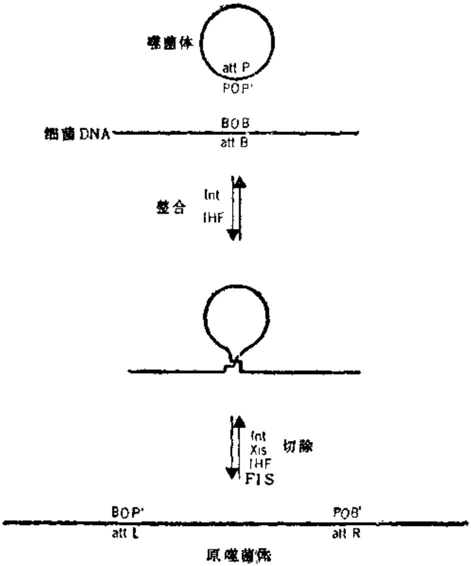



图 11-17 λ 噬菌体 DNA 整合及切除示意图

这一反应由 λ 噬菌体基因int 的产物整合酶(integrase 简写为 Int)催化。Int 只能催化 BOB′ 和 POP′ 之间的重组,不能催化 BOP′ 和 POB′ 之间的重组。因此在只有整合酶存在时,上述的反应是不可逆的。Int 是一种 DNA 结合蛋白质,对 POP′ 序列有强烈的亲合力,同时它具有 I 类拓扑异构酶活性。整合反应还需要一种叫做整合寄主因子 (integartion host factor, 简称 IHF)的蛋白质,该蛋白质含有两个亚基,均由寄主基因编码,其中的一个是受 SOS反应控制的基因 *himA*。IHF 也与 *att* 位点结合。

切除反应发生在原噬菌体两端的 *attL*(BOP′)和 *attR*(POB′)之间,由此产生λ噬菌体环状 DNA 和细菌 DNA。重组后 λ 噬菌体的 *att* 位点恢复为 *attP* (POP′),细菌 DNA 的 *att*位点恢复为 BOB′。可以把切除反应写成:

$$\underset{\text{原噬菌体}}{\mathrm{BOP}' + \mathrm{POB}'} \xrightarrow[\mathrm{IHF(FIS)}]{\mathrm{Int\ Xis}} \underset{\text{细菌}}{\mathrm{BOB}'} + \underset{\text{噬菌体}}{\mathrm{POP}'}$$

催化这一反应的蛋白质因子除了 Int 和IHF之外,还需要一种叫做切除酶(excisionase, 简称 Xis)的蛋白质,由 λ 噬菌体的 *xis* 基因编码。Xis 与 Int 结合形成复合体,该复合体具有与 BOP′ 和 POB′ 结合的能力,促使二者之间的相互作用和重组,Xis-Int 复合体不能催化 BOB′ 和 POP′ 之间的重组。故在 Xis 大量存在时,切除作用是不可逆的。此外,寄主基因编码的另一个蛋白质逆转刺激因子 FIS (factor of inversion stimulation)亦可能参与 λ 原噬菌体的切除反应。

λ 噬菌体的整合与切除受到严格的遗传学控制。λ DNA 侵入细胞后,整合能否发生取

决于 Int 蛋白质的合成。*int* 基因的转录调控和 *cI* 基因(编码 λ 阻遏蛋白)的调控是一致的。当 CⅡ 蛋白质大量存在时,它可以作为正向调节因子促进 λ 阻遏蛋白的产生;同时与 *int* 基因的启动子 $P_I$ 结合,促进 RNA 聚合酶从这里起始转录产生 Int 蛋白质(图 11-18)。$P_I$ 位于 *xis* 基因中,CⅡ 蛋白质与 $P_I$ 结合,使 *xis* 基因失活,这保证了在溶源化过程中 Int 蛋白质起作用时没有 Xis 蛋白质存在。否则,刚刚整合的 λ DNA 会马上被切除下来。λ 原噬菌体的切除需要 *xis* 和 *int* 基因同时转录产生 Xis 和 Int 蛋白质。当寄主细胞内出现 SOS 反应时,RecA 蛋白质促进了 λ 阻遏蛋白的水解,把 $O_L$ 和 $O_R$ 从阻遏状态解放出来。由 $P_L$ 处发动的转录使 *xis* 和 *int* 基因表达产生 Xis 和 Int。讲到这里,读者必定要问 λ 噬菌体的返回调控机制(见第八章第七节)为什么又不起作用了呢?原来,λ 噬菌体在整合之后,其 *sib* 位点与整合酶基因之间相距 $4.2\times10^6$ 碱基对了。这样,*sib* 位点就不可能出现在左向转录物中,也就不存在返回调控这回事了。

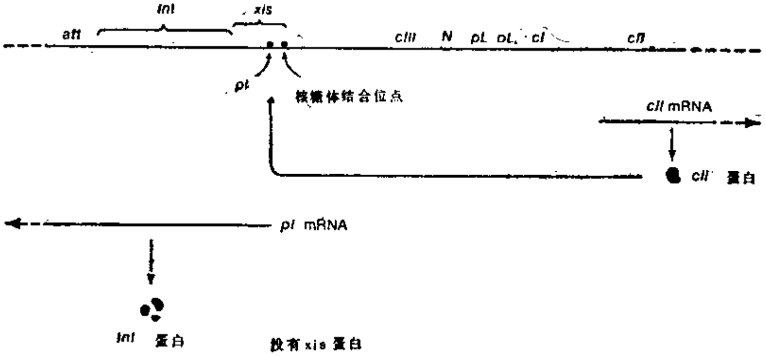



图 11-18　溶源化时 *int* 基因和 *xis* 基因的调控

## 二、λ 噬菌体整合的分子机制

上面谈到,λ DNA 的整合是一种位点特异性重组过程。重组发生在作为 *attB* 和 *attP* 各自核心的 O 序列中,因此 O 序列又叫做核心序列(core sequence)。核心序列在 *attB* 和 *attP* 以及由它们形成的 *attL* 和 *attR* 中均完全一致。它全长 15bp,富含 A-T 碱基对,无碱基倒转对称性(图 11-19)。

λ 的整合涉及到 *attB* 和 *attP* 的核心序列中链的断裂与复合(图 11-19)。断裂在双链的不同位置上发生,形成参差不齐的 5′ 单链末端,类似于某些限制性内切酶形成的粘性末端。这种参差断裂形成了 5′-OH 和 3′-P 的切口。5′ 端单链区全长 7 个碱基。同位素标记试验证明,两个核心序列中参差断裂完全相同,复合过程不需要任何新的 DNA 合成。

*attB* 和 *attP* 核心序列两侧 B、B′、P 和 P′ 序列的界限可以通过在这些序列中制造位置不同及长短不同的缺失并分析它们对重组的影响而鉴定。结果表明,attP 中 P 序列的上限为 −152(以核心序列的中心碱基为零位),P′ 序列的下限为 +82,整个 *attP* 共长 235bp。*attB* 则短得多,只有 23bp 长,其上、下限分别为 −11～+11,也就是说,B 和 B′ 实际上只分别包括

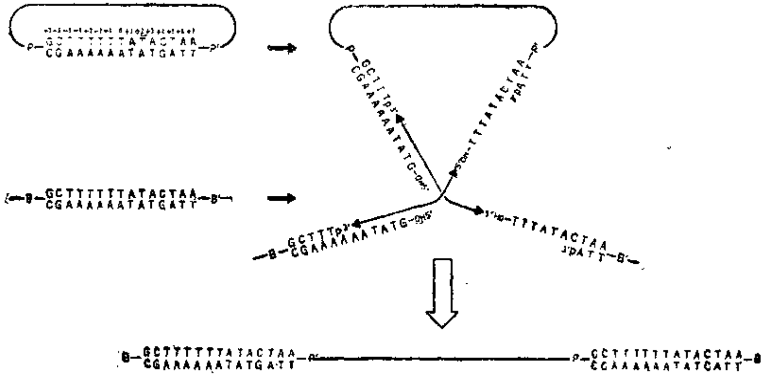


图 11-19 *attB* 和 *attP* 中相同核心序列在整合过程中的参差断裂及复合，由此形成 *attL* 和 *attR*

核心序列上、下游各 4 个 bp。*attB* 和 *attP* 的不同长度表明它们在重组中有着不同的功能。

离体条件下，Int 和 IHF 可以催化 λDNA 和寄主 DNA 的位点特异性重组。当采用超螺旋 DNA 分子作为反应底物时，几乎所有的超螺旋都依然保留在生成的整合产物中，说明整个反应中没有可以自由旋转的游离末端出现，断裂-复合的机制可能类似于 DNA 拓扑异构酶 I 催化的反应机制，不同之处仅仅在于位点特异性重组中复合发生在两条不同的DNA分子上的断口之间。Int 具有拓扑异构酶 I 活性，它可能直接参与了断裂-复合反应。

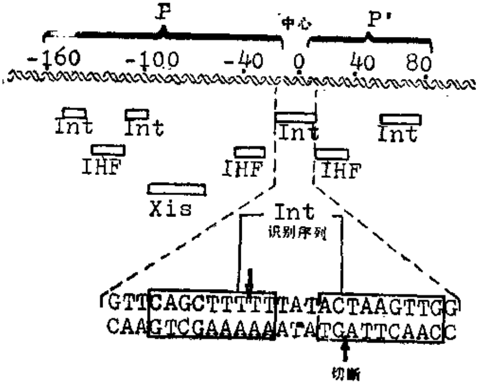



图 11-20 *attP* 位点上 Int 和 IHF 的结合位点以及 Int 在核心序列处的识别位点和切点

离体条件下，每生成一个重组 DNA 分子需要大约 20～40 分子的 Int 蛋白质和大约 70 分子的 IHF。这种化学剂量关系说明 Int 和IHF 蛋白质的主要功能是结构性的而不是催化性的。实验表明，这两种蛋白质与 *att*位点的特异性序列结合。图11-20所示为*attP*中的各结合点，其中 Int 蛋白质的结合点有 4 个，即包括核心序列在内的一段30bp序列，P′ 中的一段 30bp 序列，p 中的两个 15bp 序列。结合点的不同长度很可能反映了参与结合的Int 蛋白质单体的不同数目。*attP* 中 IHF 的结合点有 3 个，每个约长 20bp。IHF 的结合点和 Int 的结合点彼此靠得很近，它们加在一起占据了 *attP* 区域的大部分。*attB* 位点中只有一个 Int 结合位点，长约 15bp，位于核心序列中。在 *attP*（也可能在 *attB* 中）的核心序列处，Int 的识别位点相对于核心序列中心点来说是不对称的，略偏向右边，这使断裂切点的中心偏向于右边 1bp（图 11-20）。Int 和 IHF 与 P 和 P′ 序列中特异性位点结合的功能尚不清楚。

## 第六节 细菌中的转座成分[27~34]

### 一、转座成分概述

50年代初期，McClintock 通过对玉米遗传现象的细微研究，首先发现了能在基因组内不同区域转移的控制成分(controlling elements)。随后，可转移遗传成分在果蝇中发现。不过，由于人们的认识和分析技术的局限性，这些发现在当时未引起足够的重视，(从 McClintock 1951 年发表她的发现到 1983 年获得诺贝尔奖金，相隔达 32 年之久)。到 60 年代后期人们认识到大肠杆菌中某些突变是在染色体上插入了一长段序列所致，对可转移成分遗传学研究才迅速开展起来。短短十几年中，许多可转移成分已在细菌中和各类真核生物中发现，对它们的组成，转移机制，在基因交流和进化中的意义有了初步的认识。这些可转移成分现在一般叫做转座单元，简称转座元(transposons)，也叫做转座子。转座元的转移过程叫转座(transposition)。本节将介绍细菌中的转座元。真核生物中的转座元及与转座作用有关的一些遗传学现象，将在第七节至第九节中叙述。

需要指出的是，"转座"一词并非十分切意。因为转座作用涉及到转座元的复制，当一个拷贝插入基因组的新位置(受点或靶点)后，原来位置(供点)上仍然保留一个拷贝。因此转座作用并不是转座元由基因组的一处转移到另一处的简单过程。

细菌中所有的转座元都携带有其转座所必需的基因。某些"寄主"功能如 DNA 聚合酶、DNA 旋转酶等也是转座所需要的。和前面介绍的同源重组和位点特异性重组不同，转座作用不依赖转座元和靶点之间以及供点和靶点之间任何的序列同源性。

转座元插入后，使靶点处的基因失活。而且，转座元的存在及复制转座加重了细胞的生理负担，因此具有一定的选择负势。但另一方面，转座元也具有十分重要的进化意义。首先，转座作用可以造成直接的或间接的基因重排。直接效应包括转座引起的复制元融合，基因缺失，倒位或易位；间接效应指插入在不同染色体上或同一染色体不同位置上的同一种转座元可作为袖珍同源区域而允许重组发生，从而导致缺失、插入、基因扩增、倒位及易位。其次，转座作用是某些细菌质粒进化的重要方式。自然界的许多质粒的基因成份可能是在长期进化中通过转座作用逐步积累在一起的结果；某些质粒如大肠杆菌的 F 因子在细菌染色体上的整合常常是通过质粒上的转座元和染色体上相同的转座元之间的相互重组而实现的。第三，在有些情况下，细胞基因组或噬菌体基因组能够"捕捉"一些转座元，使之变成基因组中的一个组分，负责具有某种调节功能的特异性转座作用。这方面的例子有鼠沙门氏菌鞭毛相变中的 995bp 可倒位区域和大肠杆菌噬菌体 Mu 中的 G 片段。

就转座元本身而言，它们的可转移性为自然选择提供了无数目标。有的意见认为，某些转座元是自私的 DNA (selfish DNA)，其功能只是为了自身的扩增、繁衍。这些转座元同基因组的关系尤如寄生物与寄主的关系，它们的存在数目和密度反映了其潜在的巨大扩散能力和因此给细胞带来的危害之间的平衡。

细菌中的转座元最初由它们对某些基因的插入失活效应而发现。插入或是妨碍了转录，或是妨碍了翻译；而且插入往往表现出极性效应；转座元在操纵元上游基因的插入影响到下游基因的表达。极性效应的原因之一是转座元序列中有终止子信号，它们的插入会造成 mRNA 转录的终止。另一个原因是转座元序列中含有无义密码子，造成翻译的终止，由此影响到后续

基因的翻译。由于转座元有自己的基因及其调控信号，它们的插入在一些情况下能够改变靶点下游基因的表达状况，鼠沙门氏菌鞭毛相变中基因表达的转变就是一例。

绝大多数转座元都可以在染色体的任何碱基序列内插入。但其中有些程度不同地倾向于在某些特定的靶点序列插入，这种情况类似于重组热点。有的转座元倾向于优先插入染色体上的某些较大区段内，这些区段可长达3Kb左右，而在这些区段内插入则是完全随机的。这种对插入靶点的选择方式叫做"区域性优先"(regional preference)方式。造成区域性优先选择方式的原因尚不清楚，不过有一点可以肯定，即它不是取决于DNA的碱基序列，而是取决于其它因素如DNA的超螺旋状况或DNA-蛋白质的结合状态等。一个转座元在某一靶点上的插入常用双冒号表示。如 *galT*::Tn4 意指转座元Tn4在 *galT* 基因中的插入。

转座作用的频率和自发突变频率相似，介于每世代每细胞 $10^{-5}$～$10^{-7}$ 之间。转座元的切除频率要低得多，介于每世代每细胞 $10^{-6}$～$10^{-10}$ 之间。

细菌中的转座成分依组成结构可以分为以下几类：插入序列(insertion sequences)，复合转座元(composite transposons)，TnA转座元家族(TnA family)和可转移噬菌体(transposable phages)。其中以插入序列最为简单，故本节将首先介绍，然后介绍其它类型。

## 二、插入序列

插入序列因它们的插入使靶点处基因失活而被发现。插入序列有许多种。每一种的命名以缩写字母IS开头，后加不同数字。插入序列是细菌染色体、质粒和某些噬菌体的正常组分，例如大肠杆菌某一标准菌株中含有8个拷贝的IS1，各5个拷贝的IS2和IS3等，大肠杆菌质粒F因子含有一个拷贝的IS2及2个拷贝的IS3等。P1噬菌体含有IS1，其它噬菌体也含有某些插入序列。

表11-2 细菌中一些常见的IS及类IS

| | 长 度 (bp) | 末端倒转重复 (bp) | 靶点同向重复 (bp) | 编码蛋白质的可能数目 | 靶点选择 |
|---|---|---|---|---|---|
| IS | | | | | |
| IS1 | 768 | 23 | 9 | 2 | 区域优先 |
| IS2 | 1327 | 41 | 5 | 2 | 热 点 |
| IS3 | 1428 | 18 | 11或12 | 2 | 热 点 |
| IS4 | 1195 | 16 | 4 | 4 | 热 点 |
| 类IS | | | | | |
| IS10R | 1329 | 22 | 9 | 3 | 热 点 |
| IS50R | 1531 | 9 | 9 | 3 | 热 点 |
| IS903 | 1057 | 18 | 9 | 2 | 不 详 |

表11-2上半部分为细菌中一些常见的插入序列的组成特点。每种插入序列都在两端含有短的末端倒转重复(inverted terminal repeats)，其长度大多介于15～25bp之间。就一种IS而言，两个末端倒转重复的碱基序列高度相关，但通常不完全一致。

当一个转座元(包括插入序列)转座时，基因组DNA上新的靶点序列出现倍生，使插入序

列的两侧各有一段相同的靶点序列。它通常很短，大多数介于 5～9bp 长，两个重复单位方向相同，即同向重复(direct repeats)，图 11-21 为某一假设的插入序列的末端倒转重复以及靶点序列同向重复的模式图。靶点序列本来含有一个拷贝的 $\frac{ATGCA}{TACGT}$ 序列。转座发生后，这个序列出现倍生，使转座子的左右两侧各存在一个拷贝。如果在染色体上某一 DNA区段中发现有如图 11-21 的序列组成，即一段序列的两端互为倒转重复。其外侧各自紧连着一段短些的

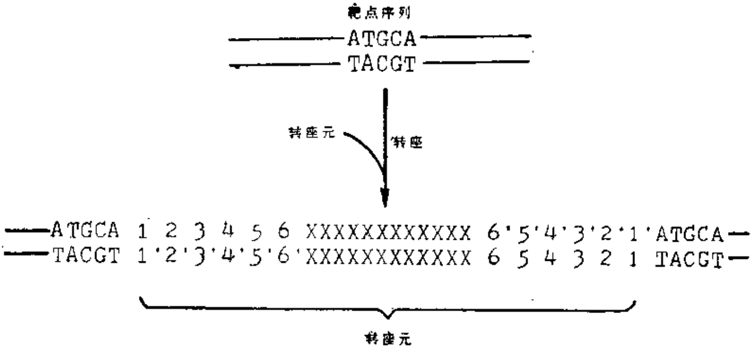



图 11-21 插入序列(及许多其它类型的转座元)末端倒转重复及靶点同向重复示意图

同向重复序列，则这很可能是一个转座元。对于大多数转座元来说，每次转座事件造成的同向重复序列各不相同，但序列的长度是一定的。与此相反，同一种转座元不同拷贝的末端倒转重复序列总是一样的。末端倒转重复非常重要，其中许多突变都会妨碍该转座元的转座作用。这些末端突变的作用是顺式的。

插入序列只含有少数与自身的转座有关的基因。最短的 IS1 含有两个方向相同的开放阅读框，编码 InsA 和 InsB 两个多肽链。其余的插入序列如 IS2 和 IS4 等含有单一的开放阅读框，自一端的末端倒转重复单位内侧始，至另一端的末端倒转重复单位止。这个长的单一开放阅读框的蛋白质产物已在有些插入序列中分离到，它们具有转座酶 (transposase) 活性。除此之外，IS2 和 IS4 等还在反义链上含有一个或几个编码区。目前尚不清楚它们是否真的转录翻译产生蛋白质产物。表 11-2 中“编码蛋白质的可能数目”一栏包括了反义链上的这些编码区。

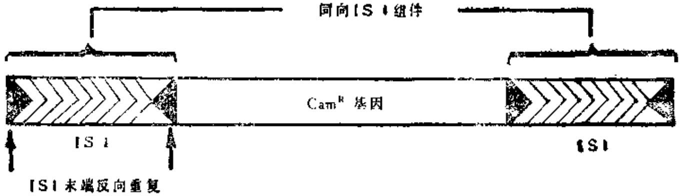
### 三、复合转座元

复合转座元是较复杂的一类转座成分。除了与自身转座有关的基因外，它们还含有其它基因如抗药性基因等。这些转座元的命名以 Tn起头，后面加上数字表示不同的转座元。除了复合转座元外，TnA 类型的转座元也含有其它基因，其命名法也是以 Tn 起头，后加数字；不过TnA 类转座元基本结构不同于复合转座元，这将在本节“六”中论述。

复合转座元含有一个中心序列和位于其两侧的臂 (arm)。中心序列含有抗药性基因等遗传信息。左、右两个臂在序列上相似，因此有时又叫长末端重复(lengthy terminal repeats)。细菌中，狭义上的转座元只包括具有上述结构的复合转座元。

有些复合转座元两条臂为相同的插入序列，例如 Tn9 的臂就是两个序列方向相同的 IS1 (图 11-22a)。不过臂的方向只在少数情况下相同，在大多数场合下，臂的方向是相反的，表现为倒转重复。如 Tn1681 的臂就是方向相反的 IS1 (图 11-22b)。但是由于每一个插入序列都有自己的末端倒转重复，所以无论臂的方向是相同还是相反，两边臂中最外侧序列总是表现为末端倒转重复。

(a)




(b)

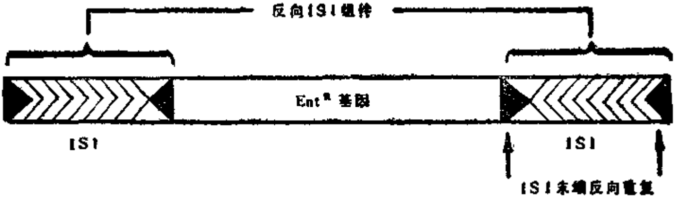



图 11-22　复合转座元 Tn9(a)和 Tn1681(b)的组成

有些复合转座元的臂在结构上十分类似于目前已发现的插入序列，即中间为一个单一的开放阅读框，两侧为较短的末端倒转重复序列。不过由于至今尚未在任何基因组中发现它们的单独存在，所以称它们为类插入序列。其命名方法同插入序列，也以 IS 起头，后面的数字通常与它们所存在的转座元的 Tn 数字有关。表 11-2 下半部分为一些类插入序列的特性，可以看出，它们与表的上半部分列出的插入序列的各种特性非常相似。在复合转座元中，两侧类插入序列的方向也有相同的和相反的两种，使整个复合转座元的最外侧也总是表现为末端倒转重复。

复合转座元的臂一方面是整个转座元的组成部分，另一方面它们本身又是插入序列或类插入序列。因此这些臂常常被称为IS或类IS组件(modules)。表11-3列出了某些复合转座元组件的特点。有些复合转座元的两个组件完全相同，如 Tn9，Tn903 和 Tn1681。有的两个组件不尽相同，如 Tn10 的左、右组件 IS10L 和 IS10R 之间有 2.5% 的碱基序列差异；不过 IS10L 和 IS10R 的末端倒转重复序列是完全相同的。虽然尚未在各类基因组中发现类 IS 组件的独立存在，但其中一些在实验条件下能够使自身转座，如 IS10R 和 IS50R 就具有此能力。一般来说，当一个复合转座元的两侧组件不同时，该转座元的转座作用主要依靠其中一个组件的功

能(如Tn10中的IS10R和Tn5中的IS50R)；当两侧组件完全相同时，其中任何一个都能行使使该复合转座元转座的功能。

**表11-3 一些复合转座元的基本特征**

| 转座元 | 长度(bp) | 遗传标记 | 末端组件 | 组件排向 | 组件间关系 | 组件功能 |
|---|---|---|---|---|---|---|
| Tn10 | 9300 | $tet^R$ | IS10R<br>IS10L | 倒转 | 2.5%差异 | 完整功能<br>功能减弱 |
| Tn5 | 5700 | $kan^R$ | IS50R<br>IS50L | 倒转 | 1bp差异 | 完整功能<br>无功能 |
| Tn903 | 3100 | $kan^R$ | IS903 | 倒转 | 相同 | 两者都有功能 |
| Tn9 | 2500 | $cam^R$ | IS1 | 同向 | 相同 | 两者都有功能 |
| Tn1681 | 2086 | $ent^R$ | IS1 | 倒转 | 相同 | 两者都有功能 |

复合转座元的组件是如何起源的呢？比较流行的看法认为，某个插入序列转座时插入离给点不远的受点(靶点)。给点插入序列和受点插入序列之间的DNA区段相当于复合转座元中的中心区域，而插入序列本身则相当于两端的组件。开始时两个组件完全相同，在以后的进化中它们或是仍然保持相同，或是因为突变的积累而出现歧化。由于一个组件就足以使整个复合转座元发生转座作用。因此在分子水平上不存在使两个组件保持一致的选择压力，另一个组件会以很快的速度积累变异。

两个靠得不太远的插入序列能够使它们之间的任何序列以及它们自身发生转座，图11-23是对这种情况的说明。如果Tn10(携带抗四环素基因 $tet^R$)处在一个不大的环状质粒中，则既可以把它的两个组件看作是Tn10本身的组件，又可以把它们看作是环状DNA除了Tn10之外的其余部分的两端组件。由这两个组件起的转座作用或是包括Tn10原来的中心区域，或是以环状DNA中Tn10以外的序列作为中心区域。在这两个转座事件中两个组件的方向总是反的，但相对于中心序列而言相反的形式又不相同。这说明IS10R无论采取何种方向都具有活跃的转座功能。在本例中以Tn10以外的序列作为中心区域的转座频率与该区域长度成负相关，序列越长(即两个组件之间的距离越长)，频率就越低。

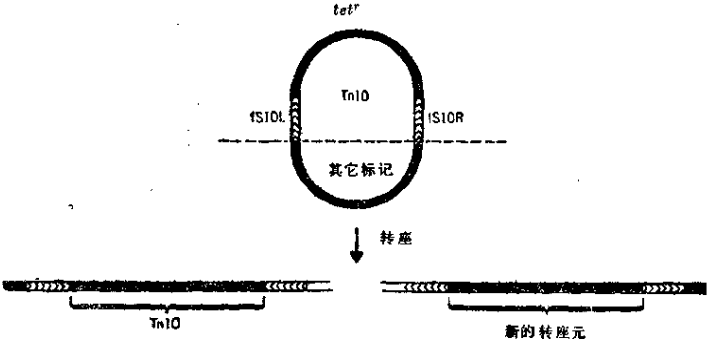



图11-23 两个末端组件可以使介于它们之间的任何序列发生转座

## 四、Tn10 和 Tn5 的末端组件与转座的调节

上面谈到某些复合转座元的两个组件不尽相同。其中只有一个行使功能，这些转座元中研究得最多的是 Tn10 和 Tn5，以下将分别讨论。

1. Tn10　Tn10 两端的末端倒转重复有 22bp 长。但只有最靠外侧的 13bp 是转座作用所必需的，改变这段序列的突变使转座作用丧失。这些突变的效应表现为顺式，与它并存于同一细胞中的野生型末端组件不能恢复突变体的转座能力。

对大量 Tn10 两侧靶点同向重复序列的分析表明，Tn10 的转座作用表现为一定的热点效应，需要一段 9bp 长的靶点序列，即 $\frac{\text{NGCTNAGCN}}{\text{NCGANTCGN}}$。一个靶点越是接近这一序列，它的转座热点效应就越强。

Tn10 的 IS10R 组件提供了该复合转座元的绝大部分转座活性。IS10R 因突变而失活后，IS10L 所表现的活性只是 IS10R 的 1～10%。IS10R 的结构如图 11-24。它拥有一个编码 74KDa 蛋白质的必需基因；反义链上还有一个重叠的阅读框架，其功能尚不清楚。上述必需蛋白质水平可以被一些所谓"跃高"(high-hopper)突变所增进，转座频率同时提高。因此该蛋白质的含量可能是限制转座频率的主要因素。此蛋白质因突变而失活后，转座作用也跟着丧失。这种缺陷可以反式地被细胞内并存的另一个野生型 Tn10 互补校正；但校正后，转座频率远不如野生型 Tn10 的高。这就说明此蛋白质有一定的"顺式"效应，它优先发起自己的模板转座元的转座。这种情况普遍存在于以插入序列为媒介的转座作用之中。

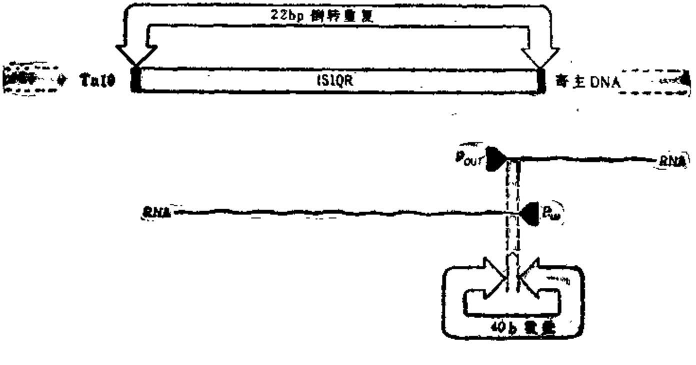



图 11-24　IS10R 中的两个启动子，其中 $P_{OUT}$ 比 $P_{IN}$ 强。

IS10R 中靠近外侧末端不远的地方存在着两个启动子(图 11-24)，一个叫做 $P_{IN}$，负责起始转录 IS10R 中的基因；另一个叫做 $P_{OUT}$，由此处起始的转录穿过末端倒转重复单位和"寄主"DNA 中的同向重复单位并向下游寄主 DNA 转录而去。此一行动很可能会激活紧靠 Tn10 右方的寄主基因。IS10R 的基因表达受所谓"多拷贝抑制"(multicopy inhibition)作用的调控：一个细菌细胞中 Tn10 的转座能力可以通过向该细胞中导入众多拷贝的 IS10R 而大大降低。多拷贝抑制需要 $P_{OUT}$ 序列的存在，并且只有在 IS10R 基因是从 $P_{IN}$ 处而不是从其它地

方起始转录时才发生。不过有趣的是，多拷贝抑制效应是在翻译水平上而不是在转录水平上：$P_{IN}$ 和 $P_{OUT}$ 的转录模板序列有 40bp 的重叠。由于 $P_{OUT}$ 的启动效应比 $P_{IN}$ 的强得多，由前者起始转录产生的外向 RNA 的量也比由后者起始转录产生的 IS10R 基因 mRNA 的量多得多；细胞中大量的外向 RNA 和 IS10R 基因 mRNA 在重叠区域的碱基配对妨碍了 IS10R 必需蛋白质的合成。

2. Tn5　Tn5 中的 IS50R 行使该转座元的全部转座功能，它比毫无此功能的 IS50L 只相差一个碱基对。这个碱基对在 IS50R 中是正常的氨基酸密码子组分，而在 IS50L 中却构成 Ochre 无义密码子。也就是说若把 IS50R 看作野生型序列，则相差的那个碱基对在 IS50L 中表现为无义突变。这一碱基代换的效应见图 11-25。IS50R 中含有一个阅读框，可以利用它产生两种蛋白质：除了蛋白质 1 比蛋白质 2 在 N-端多 40 个氨基酸外，二者的其余部分完全相同，这两种蛋白质合成中转录和翻译的调控机制尚不十分清楚，但有一点很清楚，即蛋白质 2 的合成速度比蛋白质 1 的高。IS50L 中的 Ochre 碱基代换一方面妨碍了蛋白质 1 和 2 的翻译，使合成的多肽产物缩短一截，不再具有发起转座的能力；另一方面它的出现也形成了一个能够起始 Tn5 中心区域内新霉素磷酸转移酶 Ⅱ 基因转录的启动子。该酶的存在为细胞提供了抵抗新霉素及卡那霉素等抗生素的能力。因此这一碱基代换是 Tn5 转座元表达其抗药性功能所必需的。

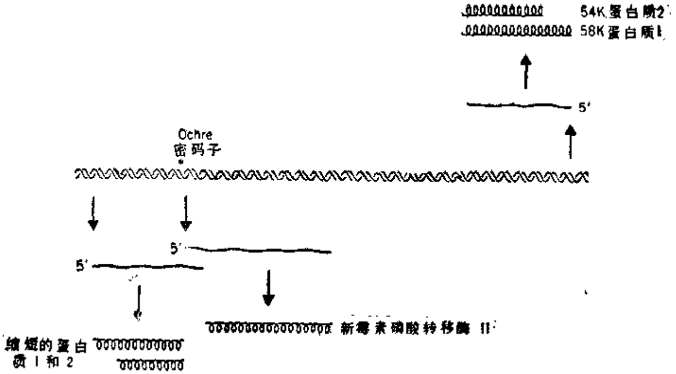



图 11-25　Tn5 中 IS50L 的一对碱基差异使蛋白质 1 和 2 不再合成，但同时却制造了促使新霉素磷酸转移酶基因表达的启动子。

蛋白质 1 和 2 的功能相关但不相同。蛋白质 1 是 IS50R 以及整个 Tn5 的转座作用所必需的。野生型的蛋白质 1 对缺陷型 IS50R 组件的反式互补作用极弱，对于转座功能表现出很强的顺式作用。蛋白质 2 是转座作用的抑制因子，它的作用是反式的。但它的功能不是妨碍某些基因的表达而是妨碍转座过程中某一尚未查明的步骤的通过。一种可能性是，蛋白质 2 具有和蛋白质 1 同样的 DNA(或蛋白质)结合位点，它可以通过和蛋白质 1 竞争这些结合位点而抑制后者的转座功能；另一种可能性是蛋白质 2 通过和蛋白质 1 之间形成某种无转座活性的寡聚物而抑制蛋白质 1 的功能。

Tn5 进入一个新的寄主细菌细胞后的开始阶段，细胞里蛋白质 2 的浓度很低，Tn5 能够

以较高频率发生转座。一旦它在细胞内"站稳脚跟"之后，转座频率随之下降，而且这时若有新的 Tn5 进入细胞，它们的转座能力也受到抑制。造成这种情况的原因是，随着开始进入的 Tn5 稳定存在状况的建立，蛋白质 2 的浓度也逐渐升高，对任何新的 Tn5 转座事件的抑制能力也跟着提高。由于促进 Tn5 的转座的蛋白质 1 的作用主要是顺式的，无积累效应，当然不能和反式的具积累效应的蛋白质 2 的竞争性抑制作用相抗衡。

比较调节 Tn10 转座的 $P_{OUT}$-$P_{IN}$ 系统和调节 Tn5 转座的蛋白质 1-蛋白质 2 系统，可以看出二者在形式上的明显差异；但另一方面它们有一个共同的特点，那就是限制转座频率。这种限制效应使寄主细胞免于因过多转座元的存在而丧生。因而从总体上看这是转座元在长期进化过程中形成的一种更好地保存自己和扩增繁衍的方式。

### 五、转座的机制

所有转座元转座的机制都很相似，转座作用包括两步主要反应。即转座元本身的复制和基因组靶点序列的断裂以及随后的倍生。整个转座过程中没有游离转座元分子产生。

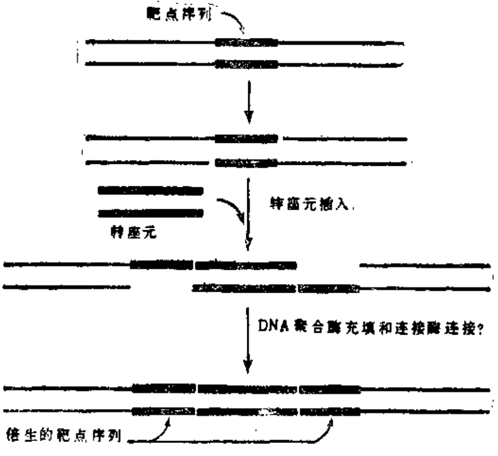



图 11-26　靶点序列同向重复产生的可能途径

靶序列的断裂和倍生造成转座元的插入以及两侧同向重复单位的生成。其大致途径见图 11-26。一种 DNA 内切酶活性通过机制尚不清楚的方式随机地或有一定优先性地选择基因组 DNA 上的一个靶点序列并在该序列两侧各一条单链上造成一个切口，切口之间的距离决定了将来转座元两侧同向重复单位的长度。然后转座元插入参差不齐的切口之间，二者共价连接起来；由此形成的两段靶点序列单链区域由 DNA 聚合酶 I 或其它类似的酶活性填补，再由连接酶将断口连接起来。这样就完成了倍生过程。上述反应中，没有转座元和靶点序列之间碱基配对的发生，也不需要二者之间序列同源性的反应步骤。不过应注意，这里强调的是靶点序列的断裂和倍生，实际上转座元的插入要比图 11-26 中所示的复杂，详情见下述。

转座元复制转座的机制过去一直是个谜。近来通过对以转座元为媒介的共合现象(coin-

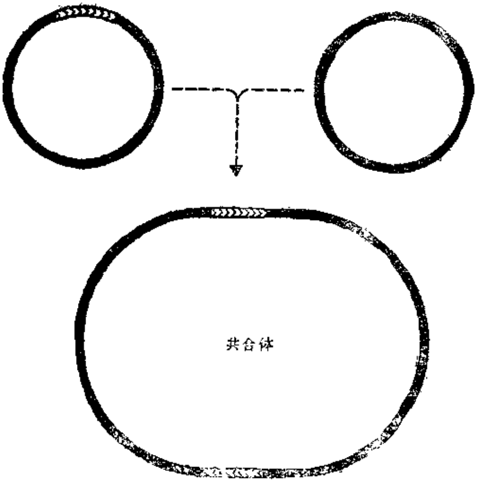



图 11-27　转座中复制元融合形成共合体，共合体中含有两个转座元

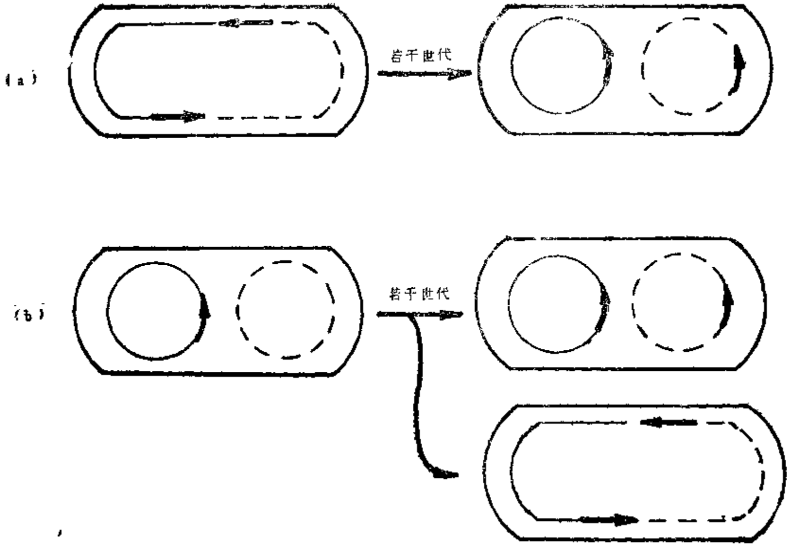



图 11-28　共合体的形成与转座作用

tegration)的研究，已经开始有了一点眉目，在此基础上提出了一些旨在解释转座过程的模型。

当含有一个转座元的复制元和另一个复制元处在同一个细胞中时，二者可以融合在一起，形成共合体(cointegrate)。这一过程有时叫做复制元融合 (replicon fusion)(图 11-27)。参与融合的复制元可以是质粒、噬菌体或细菌染色体。不过，所形成的共合体不只含有一个转座元，而是同时含有两个；它们分别处在两个复制元的两个交界处，具有相同的方向。由复合转座元所导致产生的共合体中新的转座元拷贝可能是整个的复合转座元(该共合体叫做Tn共合体)，也可能是其中的一个 IS 组件(该共合体叫做 IS 共合体)。所有的转座元都可以导致共合体的产生，但共合频率有高有低，共合体的稳定性也各有差异。

如果一个细菌细胞中含有如上介绍的那样由两种质粒形成的共合体，在繁殖了许多世代以后，一些后代细胞中的共合体便会消失；与此同时原来构成共合体的两种质粒又单独存在，但此时每种质粒中均含有一个转座元的拷贝(图 11-28a)。如果原来细胞中含有的不是共合体而是两种质粒，其中的一个含有一个转座元，那么在繁衍了许多世代之后，一些后代细胞中出现共合体，而另一些后代细胞中这两种质粒仍单独存在，但都含有一个转座元拷贝(图 11-28b)。这些结果表明，共合体很可能是某些转座元转座过程的中间物。共合体的拆分导致了原来两种质粒的重新单独出现，转座作用最后完成。有些转座元如 Tn5 可以导致非常稳定的共合体的形成，这些共合体可能是转座的一种最终结果而不仅仅是转座中间物。

目前有许多旨在解释转座过程分子机制的模型，其中较为大多数人接受的模型如图11-29所示。假设给体分子和受体分子均为环状 DNA 分子。首先在给体分子转座元两侧各一条单链上及受体分子靶点序列两侧的各一条单链上产生切口。然后，受体分子靶序列突出的单链末端和给体分子转座元的切口末端连接起来，产生形如 X 的连锁分子结构。以上反应由转座元自身编码的转座酶活性催化。X 形结构中的转座元序列两边各形成一个类似复制叉的分叉，可以从这些地方开始利用转座元为模板以半保留复制的方式合成整个转座元。复制过程大概由寄主酶活性催化。复制完成后，X 形结构转变成共合体，其中含有两个方向相同的转座元拷贝。最后，再通过在内部拆分位点上(res)这两个拷贝的转座元之间发生重组，重新产生各含有一个转座元拷贝的原来的给体 DNA 分子和受体 DNA 分子。催化这步反应的酶活性叫拆分酶(resolvase)，至少在许多已知的例子中拆分酶由转座元编码。上述模型同样适用于当给体分子和受体分子之一或两者都是线性 DNA 时的转座作用。

除上述模型外，另一种比较流行的模型叫做非对称模型(asymmetric model)。它与上述模型的不同主要有两点：第一，转座元两侧的切口不是同时产生的，而是分步产生；第二，共合体不是转座过程中的中间物，而是转座过程所导致的两种最终结果之一。(另一种最终结果是转座)。有些转座元(如上面提到的 Tn5)能够导致稳定共合体的形成，这是与非对称模型一致的。但它难以解释 Tn3 共合体的拆分。目前由于尚无更多的实验证据支持非对称模型，所以这里不再详细讨论它的每一步反应。

### 六、Tn3 及其转座作用

在本节开始介绍转座元的分类时曾提到 TnA 转座元家族。它包括许多彼此相关的转座元。这类转座元较长，约 5000bp，比插入序列长得多。和复合转座元相同，TnA 家族也同时携带有负责自身的转座的基因和诸如抗药性基因等其它基因；不过和复合转座元不同的是，TnA

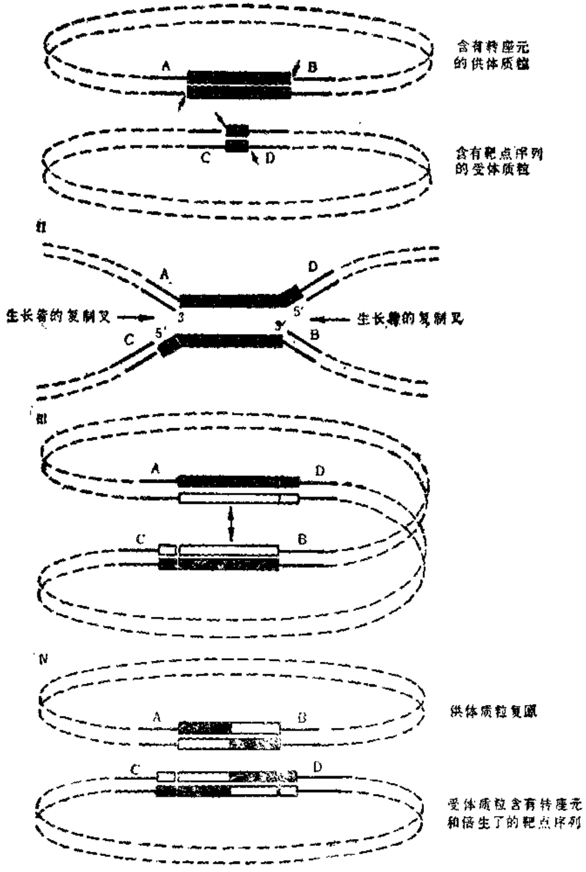



图 11-29　转座作用的一种模型

家族转座元的两端没有 IS 或类 IS 组件，有的只是长度为 37～38bp 的末端倒转重复序列。表 11-4 列出了一些TnA家族转座元的一般特性。前文中所讨论的转座元转座机制在原则上同样适用于 TnA 家族；另一方面，TnA 家族的转座作用也有一些独特之处。近年来对 TnA 家族尤其是 Tn3 的深入研究，为揭示这类转座元的转座机制和调控提供了不少新的认识。

TnA 类转座元的末端倒转重复一般长 37～38 bp，同一转座元的两个末端重复的序列通常十分接近，其中任何一个的缺失都会顺式地妨碍转座功能。转座元两侧的靶点同向重复长 5bp，这些转座元对靶点的选择有区域优先性，但在一个优先区域内的的选择是随机的。

表 11-4 一些 TnA 家族转座元的基本特征

| 转座元 | 长 度(bp) | 遗传标记 | 末端倒转重复(bp) | 两侧靶点同向重复(bp) | 靶点选择 |
|---|---|---|---|---|---|
| Tn3 | 4957 | $amp^R$ | 38 | 5 | 区域性优先 |
| Tn1 | 5000 | $amp^R$ | ? | 5 | 区域性优先 |
| $\gamma^\delta$ | 5800 | ? | 37 | 5 | 区域性优先 |
| Tn501 | 8200 | $Hg^R$ | 38 | 5 | 不 详 |

TnA 类转座元的转座过程分两个阶段。第一阶段形成共合体；第二阶段共合体拆分。每一阶段各有一个由转座元基因编码的酶催化。其中 *tnpA* 基因编码转座酶，*tnpR* 编码拆分酶。互补试验表明，这两个基因的作用是反式的。这一点和插入序列及复合转座元中转座酶的强烈顺式作用大不一样。

Tn3 转座元中，*tnpA* 基因编码的转座酶由 1 021 个氨基酸组成，分子量约为 120 KDa。该酶能够与单链 DNA 结合，它是识别并切割转座元的两端以及在靶点序列上造成 5bp 的参差不齐单链末端所必需的。虽然 Tn3 和另一个相关的转座元 $\gamma^\delta$ 末端倒转重复序列的 38 个 bp 中有 27 个相同，但它们的转座酶只能有效地催化各自的转座。这说明转座酶对末端倒转重复序列的识别是高度特异性的。

*tnpR* 基因编码含有 185 个氨基酸的(23KDa)蛋白质，它除了具有上面提到的拆分酶活性外，还具有调控 *tnpA* 基因及自身基因转录的阻遏蛋白活性，*tnpA* 和 *tnpR* 基因之间间隔 163 bp 富含 A-T 的调控区域，这两个基因的转录从这里起始，朝相反的方向进行(图11-30)。*tnpR* 的蛋白质产物正是通过与这段调控区域的结合而实现对 *tnpA* 和 *tnpR* 转录的阻遏。使 *tnpR* 基因产物的阻遏蛋白功能失活的突变可以提高转座事件的频率。不过若突变破坏了 TnpR 拆分酶的活性，则细胞内只形成共合体而无法使之拆分。这就说明共合体形成及其拆分是TnA 家族转座过程中的必然中间步骤。

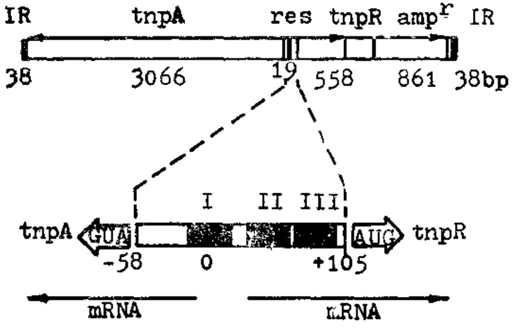



图 11-30 Tn3 及类似转座元的细微结构

TnpR蛋白质对共合体的拆分涉及到两个转座元拷贝之间的重组。它只能在转座元的一个特定位点发生，这个位点过去叫做IRS(internal resolution sequence，即内部拆分序列)，现

在改称做 *res*。当 *res* 位点缺失后，TnpR 蛋白质不再能识别共合体并行使拆分功能，从而造成共合体分子积累，转座作用无法最后完成。虽然在这种情况下寄主细胞中以 RecA 为媒介的同源重组系统可以在两个转座元拷贝之间的任何部位发生同源重组而代理拆分，但其效率毕竟要比 TnpR 在 *res* 处的特异性重组拆分的效率低得多。

TnpR 蛋白质是通过与 res 位点的结合而行使拆分酶功能的。如图 11-30 所示，Tn3 转座元上 *res* 位点处在 *tnpR* 和 *tnpA* 基因之间的调控区域内，它实际上包括3个结合点，每个约长 30～40bp，TnpR 对它们的结合相互独立。结合点 I 包括遗传学上鉴定的 res 位点。它的缺失完全妨碍了以 TnpR 蛋白质为媒介的拆分。但如果光有结合点 I 存在而没有结合点 Ⅱ 或 Ⅲ，则拆分虽可发生，但效率较低。因此野生型水平的拆分效率需要 I、Ⅱ、Ⅲ 三个结合点同时被 TnpR 占据。从图 11-30 中可见，结合点 I 与 *tnpA* 基因的转录区部分重叠。结合点 Ⅱ 和 Ⅲ 与 *tnpR* 基因的转录区重叠。因此 TnpR 在这些结合点处的结合妨碍了这两个基因的转录。Tn3 上的三个结合点序列有一定的同源性。这些序列中都含有一个对称的倒转重复区。

TnpR 蛋白质介导的拆分机制是怎样的呢？很可能在共合体上一个 Tn3 转座元拷贝的三个结合点被 TnpR 结合后改变了这一区域 DNA 的拓扑学构象，使两个处于这种状况的Tn3 转座元能够彼此互相作用。完成拆分反应。离体实验表明，拆分是一种保守反应，DNA 双链的断裂和复合在没有任何 DNA 合成和能量输入的情况下完成。断裂和复合在结合点 I 中的一段对 TnA 家族来说高度保守的 19bp 中发生。两条链的断裂点相隔 2bp，由此产生参差不齐的末端。其中的 5′ 端各与一个分子 TnpR 蛋白质共价结合。即：

5′ T T A T A A 3′
3′ A A T A T T 5′

↓

5′ T T A T　　　　+　　　　TnpR—— AA 3′
3′ A A ——TnpR　　　　　　　　TATT 5′

当每个共合体分子对应于 20 分子的 TnpR 时，拆分效率最高。

Tn3 所形成的共合体的拆分与 λ 噬菌体的位点特异性重组非常相似。这种相似性表现在以下几个方面：二者都对重组位点有特异性要求，都在断裂中产生参差不齐的末端，二者的断裂-复合反应都类似于 DNA 拓扑异构酶的催化功能，TnpR 蛋白质对 res 的结合与 Int 蛋白质对 attp 的结合在原则上相似。此外若比较 Tn3 的 *res* 位点和 λ DNA *attP* 核心序列中断裂-复合处的碱基序列，可得：

*res*　G A T A A T T T A T A A T A T
*att*　G C T T T T T T A T A C T A A

15 个 bp 中有 10 个一样。综上所述，可以把 Tn3 所形成的共合体的拆分看作一种位点特异性重组过程。

### 七、作为转座成分的 Mu 噬菌体

Mu 是一种以大肠杆菌为寄主的温和噬菌体，以裂解生长和溶源生长两种交替方式之一繁衍自已。Mu 的最初发现是基于其插入大肠杆菌基因导致突变的能力。Mu噬菌体能作为转座元区域优先性地随机插入寄主 DNA 中，它的切除频率极低。Mu 的遗传图见图 11-31。它的DNA

为线状分子，游离噬菌体状态和整合状态具有相同的基因次序。Mu DNA 不含末端倒转重复序列，这是和其它转座元不同的地方。游离 Mu 噬菌体 DNA 的两端连接着一段寄主 DNA。左端的约长 100bp，右端的约长 1500bp。这两段寄主 DNA 序列在不同 MuDNA 上各不相同，它们是在噬菌体成熟阶段将整合在寄主 DNA 中的原噬菌体 Mu 切断包装时获得的。当 Mu 噬菌体再一次整合到寄主 DNA 中时这两段序列归于消失。

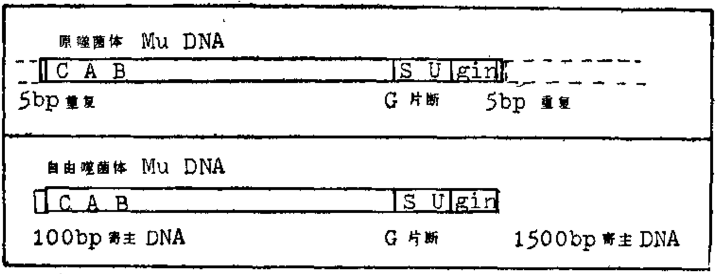



图 11-31 Mu 原噬菌体和游离噬菌体的组成；二者具有相同的基因次序

Mu 的插入有两种途径。第一，侵入的Mu可以在溶源化过程中插入寄主DNA的任意部位，造成靶点的倍生，使原噬菌体两侧各有一个 5 bp 的靶点序列重复(图 11-31)。这一点和其它转座元的转座作用相似。第二，进入裂解生长后，复制产生的后代 MuDNA 几乎全部随机插入寄主 DNA 中，一旦插入便极少再切除下来。它们能够作为转座元再造成其它靶点上的插入。由此在寄主染色体上产生越来越多的拷贝，这时的寄主染色体实际上是众多的 Mu DNA 和被它们插入肢解的寄主 DNA 在一起形成的共合体。噬菌体成熟时，共合体被切断，一个 MuDNA 分子连同它左侧的 100bp 和右侧的 1500 bp 一起被包装进入噬菌体颗粒。不处在共合体中的 Mu DNA 不能被包装。MuDNA右侧的一段3 Kb长的序列在不同分子中以相反的走向存在，这段序列叫做 G 片段，它的两端各有一个 34bp 长的倒转重复序列。图 11-31 中所示的是 G(+)走向，它倒位后则形成 G(−)走向。G 片段的倒位需要 Mu 基因 *gin* 的产物，寄主的逆转刺激因子能提高倒位的效率。

G 片段含有编码负责噬菌体吸附寄主细胞的尾丝蛋白质，G 的不同走向导致了不同的寄主特异性。当处于 G(+)走向时，S 和 U 基因表达、分别产生 56KDa 的 S 蛋白质和 21KDa 的 U 蛋白质，这类 Mu 噬菌体能够吸附大肠杆菌品系 K12 而不能吸附品系 C。当处于 G(−)走向时，S′ 和 U′ 基因表达，分别产生 48KDa 的 S′ 蛋白质和 26KDa 的 U′ 蛋白质。这类 Mu 能够吸附品系 C 而不能吸附品系 K12。无论是 G(+)走向还是G(−)走向，其蛋白质产物都需要大约 4Kb 的 DNA 编码，可 G 片段实际上只有 3Kb。这是因为 S 蛋白质和 U 蛋白质以及S′ 蛋白质和U′ 蛋白质分别由 G 片段的两条互补链所编码，它们的启动子并不存在于 G 片段中，而是存在于 G 片段左边的序列中一个共同的启动子。而且对于 S 蛋白质和 S′ 蛋白质来说，它们还有相同的一段 N 端序列，其编码区也在 G 片段左边的序列中。这样，G 片段的倒位只改变 S 蛋白质或 S′ 蛋白质的可变区($S_v$ 和 $S_v'$)以及 U 和 U′ 蛋白质的编码情况，而对于 S 或 S′ 蛋白质的相同的 N 端(Sc)却没有任何触动(图 11-32)。

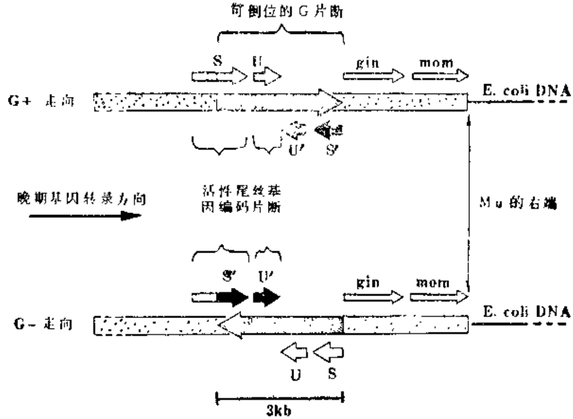



图 11-32 Mu 噬菌体的 G 片段倒位能够产生四种不同的尾丝蛋白质

#### 八、沙门氏菌鞭毛相变中的转座机制

我们在第八章第八节中谈到沙门氏菌鞭毛相变中表达调控的问题。造成Ⅰ相和Ⅱ相之间互变的原因在于 *H2* 基因和 *rh1*基因上游的一段995bp序列的倒位，其两端各有一段 14bp 的重复单位(IRL 和 IRR)。这和 Mu DNA 中通过 G 片段的倒位而改变寄主特异性的情况有相似之处，即采用倒位来调控基因表达。有趣的是负责 995bp 序列倒位的沙门氏菌 *hin* 基因的功能和负责 G 片段倒位的 Mu 噬菌体 *gin* 基因的功能在互补试验中能相互替代，它们也能和负责大肠杆菌 P1 噬菌体 C 片段倒位的 *cin* 基因的功能互相替代。同样有趣的是，虽然 Tn3 的 *tnpR* 基因功能(编码拆分酶和阻遏蛋白)不能替代 hin 的功能，但它们蛋白质产物的氨基酸序列却是很近。沙门氏菌 995bp 序列中的IRL 和IRR之间完全同源，它们及附近序列和 Mu 噬菌体 G 片段两端的 34bp 倒转重复序列也呈高度同源性。以上这些情况说明995bp序列的倒位和 G片段的倒位机制相似，都是通过类似 Tn3 的共合体拆分反应中采用的位点特异性重组的方式实现的。这些情况还表明沙门氏菌 995bp 序列和 Mu 的 G 片段有可能起源于同一种古老的转座元。在长期进化过程中，转座作用逐步受到限制，直到演化成只能倒位的目前状态。作为转座作用“遗迹”的倒位被用来调控其它基因的表达，不再与序列的扩增有关。这个假定的古老转座元可能与转座元 Tn3 的祖先有着进化上的密切关系。

## 第七节 真核生物中的转座成分(一)——真核生物转座成分的分类及其典型代表[35~45]

#### 一、真核生物中的转座成分及其分类

正如上一节开头所述，自然界第一个被发现的转座成分就是玉米中的“控制成分”。Barbara McClintock 从 1942 年到冷泉港实验室后就开始进行玉米控制成分的研究，1951年发表了她关于玉米控制成分的论文。她的开创性的工作打破了孟德尔关于基因在各个独立的

染色体上固定排列的僵化概念。此后，人们又陆续发现了许多真核生物中的可转移成分，范围从低等真核生物(酵母)一直到最高等的真核生物(人类)。因此，人们认识到，真核生物基因组的流动性比原核生物更大。

真核生物中的转座成分的转座机制,有的与细菌的转座元相似,有的则与还原病毒相似。前者的遗传信息的流动是从 DNA 到 DNA，而后者是从 DNA 到 RNA 再到 DNA。像细菌的情况一样,从DNA到 DNA 的转移过程称为转座(transposition),被转移的 DNA 单元称为转座元(transposons)。为了将 RNA 介导的转座作用与经典的 DNA 到 DNA 的转座作用相区别，于是将 RNA 介导的转座作用称为返座作用(retroposition),被转移的遗传信息单元称为返座元(retroposons)。

已发现的真核生物转座元有果蝇的 P 成分，FB 成分，玉米的 Ac-Ds 转座系统等。

真核生物的返座元的种类要复杂得多。根据其结构和组成，可以分成两个大家族(superfamily),一个是病毒大家族,另一个是非病毒大家族(表 11-5)。病毒大家族包括已发

**表 11-5 返座元及其分类**

| | | | | |
|---|---|---|---|---|
| 病毒大家族 | 还原病毒(全部) | 非病毒大家族 | RNA 聚合酶Ⅱ转录物 | |
| | 还原转座元 | | 有功能的半加工返座基因(鼠类前胰岛素原Ⅰ基因) | |
| | Ty (酵母) | | 加工的返座假基因(已发现数十种) | |
| | Copia 家族(果蝇) | | snRNA 返座假基因(人,啮齿动物) | |
| | DIRS-1 (粘菌) | | F 家族(果蝇) | LINES |
| | Bs1 (玉米) | | LINE1 家族(人,猴,鼠类) | LINES |
| | IAP (啮齿类) | | RNA 聚合酶 Ⅲ 转录物 | |
| | THE1? (人) | | 7SLRNA 返座假基因 | SINES |
| | VL30 (大鼠,小鼠) | | 未加工的(但可能截短) | |
| | | | 内部缺失的 | |
| | | | B1 家族(鼠) | |
| | | | 二聚体Ⅰ型 Alu 家族(灵长类) | |
| | | | 四聚体Ⅰ型 Alu 家族(人) | |
| | | | 7SK RNA 返座假基因 | |
| | | | 未加工的(但可能截短) | |
| | | | tRNA 返座假基因 | |
| | | | 未加工而 3 端截短的 | |
| | | | 加工的(高度突变的) | |
| | | | 单体家族(猿) | |
| | | | B2 家族(啮齿动物) | |
| | | | ID 重复序列(鼠) | |
| | | | C 重复序列(兔) | |
| | | | 复合 RNA 返座假基因 | |
| | | | tRNA+7SL (猪 AluⅡ) | |
| | | | AluI+LINE1 (人) | |
| | | | 7SK+AluI (人) | |
| | | | B2+OBY (小鼠) | |

现的所有脊椎动物的还原病毒，不管是外在的(病毒粒子)还是内在的(原病毒)，也不管是非缺陷型的还是缺陷型的。病毒大家族还包括一类结构与还原病毒的原病毒相似的转座成分，称为还原转座元(retrotransposons)，如酵母的 Ty 转座成分，果蝇的 copia 成分，人类的 THE1重复序列(?)等；有的还能形成一种没有体外阶段的类似于病毒粒子的细胞内颗粒，如小鼠的潴泡内 A 型颗粒 IAP (intracisternal A-type particles) 等。非病毒大家族的返

座元结构各异，一般都没有转座酶的编码区；有7～21bp的靶点倍生序列；正向重复序列或长或短，甚至完全没有；大多数都与成熟RNA的序列相当，而且除了四种rRNA没有相应的返座元外，几乎每一类RNA都有返座元。

本节后面一部分将叙述真核生物中一些经典的转座成分，包括转座元和病毒家族的还原转座元。第八节和第九节将分别叙述真核生物中的还原病毒和真核生物中的非病毒返座元。

## 二、酵母的Ty成分

酵母的Ty成分包括一组与非重复序列相间而散在分布的重复DNA序列(Ty这一缩写来源于transposable elements in yeast)。每个Ty成分长6.3Kb，两端各有一段334bp长的同向重复序列，叫做δ成分。每个酵母细胞基因组中约含有30～35个拷贝的Ty成分。另外还有大约100个游离的单一δ成分(solo δ elements)。Ty成分在不同酵母品系基因组中有不同的位点，这也是最初认为它们是转座成分的证据之一，Ty成分的转座可以从它们的插入使靶点处基因失活而鉴定出来。目前尚不能估计Ty成分转座的频率。一般认为它比细菌的转座频率低。

Ty成分是一大类转座成分，不同Ty成分的碱基序列可能有很大差异。最主要的两类是Ty1和Ty917。即使是同属Ty1或Ty917，不同的Ty成员之间也存在着许多差异，包括碱基代换、插入和缺失。

不同Ty成分上的δ成分也有很大的差异。但同一个Ty成分的两个δ成分则差异极小甚至完全相同。这种同一性可能归因于转座阶段或其后的基因转换，也可能归因于所出现在靶点上的Ty成分的两个δ成分均来自给点两个δ成分中的一个。Ty成分上的δ成分之间的保守性比单一δ成分之间的保守性要大些，说明δ成分可能与转座功能有关。Ty成分插入靶点后，也像细菌转座元和还原病毒那样在两侧产生靶点序列的同向重复，长度为5bp。

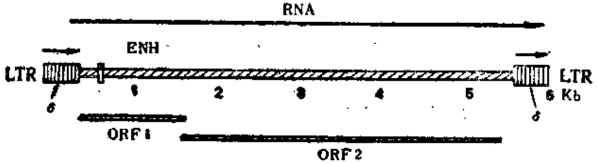



图11-33 酵母Ty1成分的大致结构(图中为Ty912)。LTR为长末端重复，即δ成分；ENH为增强子；整个Ty1成分产生一个长的mRNA，其中有两个开放阅读框(ORF1和ORF2)。

由于Ty1成分中有增强子(图11-33)，在其δ成分中又有较强的启动子，因此在Ty1成分上进行着活跃的转录活动，Ty1 RNA可以占到整个酵母细胞mRNA的5～10%。整个Ty1成分转录成一个mRNA，其中有两个开放阅读框。其中一个编码逆转录酶，而且与还原病毒的逆转录酶非常相似。可见Ty成分的转座机制完全不同于细菌的转座元。细菌转座元的转座是从DNA到DNA；而Ty成分的转座是从DNA到RNA再到DNA，类似于还原病毒的原病毒形式。实验已经证明，Ty成分的转座的确有逆转录过程。人们用基因工程的方法让Ty1成分含有一个内元，然后观察其转座情况。结果在新的基因组位点发现了这个Ty1成分，但其中的内元已经消失。因为内元只有在RNA水平上才能被拼接而去除，说明了转座过程中有RNA中间体存在。不仅Ty成分的转座机制是如此，而且现在已有足够的证据表明果蝇的

*copia* 成分的转座也是采取这种机制。

Ty 成分的存在为酵母基因组提供了许多袖珍同源区域，它们之间可以发生同源重组。如果重组在同一条染色体上发生，则造成缺失或倒位；如果在不同染色体上的 Ty 成分之间发生，则造成易位等。此外，Ty 成分可以因自身两端的同向重复 δ 成分之间的重组而圈出切除，切除后在染色体的靶点处只剩下一个 δ 成分。这也许是目前存在的许多单一 δ 成分的来源。

### 三、果蝇的 copia 等转座成分

果蝇中的转座成分最初也是通过它们的插入使靶点基因失活而被发现的。目前已发现的转座成分有好几类，其中研究得最多的是 *copia* 家族，因其众多的彼此密切相关的拷贝数而得名。此外，果蝇中还含有一些在序列上和 *copia* 不相关的转座成分，这些转座成分和 *copia* 成分加在一起占到果蝇 DNA 总量的 1% 以上，它们的特性见表 11-6。

表 11-6　果蝇中发现的几种不同的转座成分

| 转座成分 | 每基因组拷贝数 | 长　度<br>(bp) | 末端同向重复<br>(bp) | 末端倒转重复<br>(bp) | 靶点同向重复<br>(bp) |
|---|---|---|---|---|---|
| *copia* | 20～60 | 5 000 | 276 | ～10<br>(不完善倒转重复) | 3～6 |
| FB | ～30 | 500～5 000 | 无 | 250～1 250 | 9 |
| P | ～50 或 0 | 500～2 900 | 无 | 31 | 8 |

果蝇基因组中 *copia* 的拷贝数因品系而异，一般介于 20～60 之间，它们呈散在分布。不同品系的 *copia* 位点虽然有重叠，但多数不同，没有任何一个在所有品系中都存在的 *copia* 位点。两个品系的亲缘关系越近，具有相同位点 *copia* 的比例就越大，这一点可以用来估计果蝇品系的起源和 *copia* 的转座频率。通过计算，*copia* 成分的转座频率大约是每细胞世代 $10^{-3}$～$10^{-4}$。

*copia* 以及其它两种转座成分的一般结构见图 11-34。*copia* 长 5 000bp，两端具有 276bp 长的同向末端重复序列。每一个末端各有一个不完善的倒转重复序列。*copia* 成分的两侧是 3～6bp 长的靶点同向重复。*copia* 的插入无靶点序列特异性，但表现为一定的区域优先性。不同 *copia* 各成员的碱基序列差异很小，不超过 5%，其中含有许多小的缺失。

*copia* 成分两端的 276bp 同向重复序列完全相同，这和酵母 Ty 成分两端 δ 成分的情况非常相似，也同样可以用基因转换或两个重复单位来自亲体 *copia* 成分的同一个末端重复单位来解释。不过不同点是，Ty 成分的 δ 成分可以孤立存在，而 *copia* 的 276bp 重复单位不单独存在。*copia* 成分本身可以以游离环状 DNA 分子形式单独存在，这和细菌转座元的情况不同。单独存在的 *copia* 环状分子有两种，一种为 5 000 bp，另一种少一个 276 bp 末端重复单位，约为 4 700 bp。

*copia* 成分具有很高的转录活性，产生带有 poly (A) 的 mRNA。这些转录物大多被限制在细胞核中，其原因不明。现已知道，整个转录物具有一个很长的开放阅读框，编码的蛋白质与还原病毒的逆转录酶、整合酶和核酸结合蛋白具有很高的同源性。正如前面所说，*copia* 成分是一类还原转座元。

另一组转座成分 FB(foldback 的缩写)全长 500～5 000 bp，它的不同成员具有长度不同

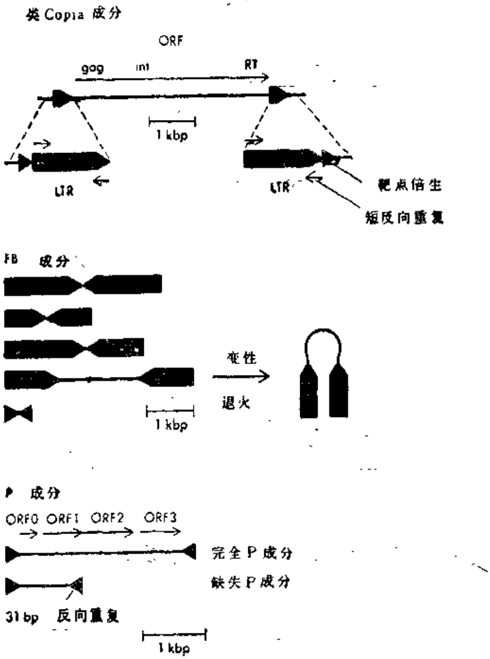



图 11-34 果蝇的 *copia* 成分，FB 成分和 P 成分的一般结构。*gag* 编码核酸结合蛋白，*int* 编码整合酶，RT 编码逆转录酶。LTR 为长末端重复。FB 成分和 P 成分都具有不同的大小。

的倒转末端重复序列，介于 250～1 250 bp 之间，同一个 FB 成分的两个末端重复序列也不完全相同。而且，介于两个末端重复序列之间的序列也长短不同，有的 FB 成分两个末端重复序列紧靠在一起，中间无间隔；有的二者之间间隔长达数千 bp。尽管如此，所有的 FB 成分末端倒转重复序列都表现一定的同源性，它们都由更小的相似重复单位组成，在结构上与卫星 DNA 相似。最末端的 30 bp 高度保守。转座时，靶点序列产生 9 bp 的倍生。已有证据表明，任何 DNA 序列，只要其两翼有 FB 成分，都可以发生转座。这种转座，往往涉及很大的 DNA 区段。显然，这种转座成分对于基因组的进化具有重要的意义。FB 成分的转座行为，非常类似于细菌中的复合转座元。在果蝇以及其他高等真核生物的基因组中，变性后的单链能产生发夹结构的 DNA 占到若干个百分数，其中真正能够转座的 FB 成分只占极少数。哪一种酶促进 FB 的转座还不清楚。

果蝇中另一种转座成分 P 也具有不定的长度。完整的 P 成分为 2 900 bp，含有四个开放阅读框，称为 ORF 0，ORF 1，ORF 2，ORF 3。内部缺失的 P 成分大小介于 500～1 400 bp 之间，保留的 ORF 各不相同。不管是完整的 P 成分还是缺失的 P 成分，两端都是完美的 31bp

倒转重复序列。现在已经知道，P 成分的四个 ORF 都是转座作用所必须的，因此只有完整的 P 成分才有转座功能。但是，内部缺失的 P 成分，只要细胞内有完整的 P 成分，也能实现转座作用。可见，四个 ORF 编码反式作用因子，其中一个是转座酶，另一个是转座的阻遏蛋白。正是由于这两种蛋白质的存在，使得有的杂交方式能产生有生殖力的后代，有的杂交方式则造成杂种不育(hybrid dysgenesis)。

当 P 品系的雄果蝇与 M 品系的雌果蝇杂交时，其后代出现杂种不育(见图 11-35)，而其他三种交配方式即 ♂(P)×♀(P)，♂(M)×♀(M)，♂(M)×♀(P)都不会出现生殖障碍。原来，两种品系中 P 成分存在情况不同，只有 P 品系中才有完整的 P 成分，而 M 品系或者缺乏 P 成分，或者只有内部缺失的 P 成分。由于卵细胞有细胞质，而精细胞则几乎完全没有细胞质。因此 P 品系的雌果蝇与任何品系的雄果蝇交配，后代都不发生生殖障碍。而M品系的雌果蝇则不同，由于其卵细胞中没有完整的 P 成分，也就没有妨碍转座的阻遏蛋白。因此，当它与 P 品系雄果蝇交配受精后，P 成分受到卵细胞质中的转录激活因子的作用而开始转录并表达出转座酶和阻遏蛋白。请回忆一下细菌转座元中的情况，许多转座酶表现出一定的顺式效应，而阻遏蛋白的作用则完全是反式的，因此受精后的短时间内，转座作用占优势。在受精卵分化过程中，生殖细胞系中的转座频率特别高，达到每细胞世代一次。如此频繁的转座会使生殖腺中的许多基因失活并造成染色体重排，最终杂交子一代表现出生殖障碍。杂种生殖障碍的程度依实际发生的 P 成分转座事件的数目而异。如果一个 F1 个体能够侥幸耐受这些转座事件而产生生殖细胞，则在第二代出现突变表型。在以后世代中，随着抑制转座酶的蛋白质因子的产生和积累，转座频率迅速下降，已经转座的 P 成分则有可能在新的靶点上被固定下来。通过杂种后代的分离，就会产生许多与其父本 P 品系具有不同 P 成分位点的新 P 品系。

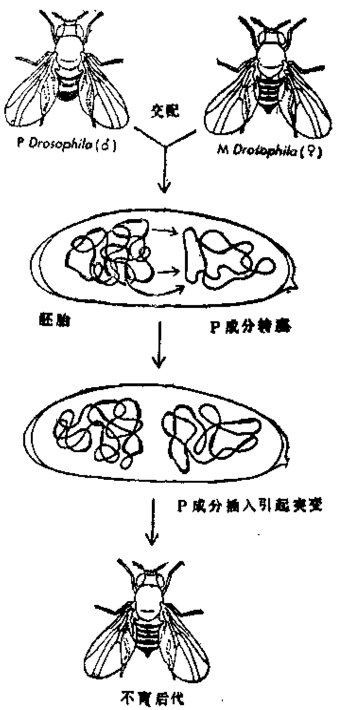



图 11-35 P 成分在果蝇胚胎中的大量转座，引起杂种生殖障碍。

有趣的是，30 多年前从野外收集的果蝇品系的后代皆为 M 型，而最近 10 年收集的则几乎皆为 P 型，这如何解释呢？一种意见认为新近的 P 型品系来自 P 成分在最近几十年中对 M 型果蝇的侵入。如果是这样，P 成分要么起源于其它物种，要么是在最近才通过某种遗传学途径产生出来。一经产生便迅速布满整个野生群体。另一种意见认为，在实验室培养条件下 P 成分的存在具有选择负势，会逐渐失去。30 多年前的果蝇品系可能本来多为 P 型的，但在长期的培养中 P 成分已消失殆尽。

P 成分造成的杂种生殖障碍可能是物种形成中生殖隔离的一种手段。如果一个物种的两个群体在地理隔离的情况下各自出现了 p 成分类的转座元，那么不同群体之间的基因交流就会受阻碍，它们之间由地理隔离继而发展为生殖隔离，最后导致新物种的形成。

**四、玉米的转座成分**

40年代，人们发现玉米细胞分裂中时常出现某些基因突变，是由一些被称为控制成分的遗传成分在基因组中变换位置而引起的。控制成分在新的靶点位置上的插入可以造成基因失活或改变附近基因的表达活性，产生出原有基因的新等位基因。由于控制成分有可能切除失去，因此这些等位基因很不稳定，故称为可变等位基因(mutable alleles)；控制成分丢失或转移他处后，不稳定的可变等位变基因即成稳定的等位基因(但不一定回复到原来的状态)。此外，控制成分的存在可以导致染色体断裂，造成附近染色体区域的缺失、重复、倒位或易位。

玉米中有好几组控制成分，如 Ac-Ds，Spm 和 Dt 等，不同玉米品系中它们的数目和插入位置很不相同。同一组控制成分的不同成员又可分为两大类，一类叫自主 (autonomous) 成分，它们具备使自身切除或转座的能力；另一类叫非自主(nonautonomous)成分，它们只有在同一组自主成分存在时才能切除或转座。互补试验表明，自主成分对非自主成分的互补效应是反式的，非自主成分大概起源于丢失了自主转座功能的成分，如各种 Ds 均来源于 Ac。

研究得较多的控制成分是 Ac-Ds 系统，其中 Ac (activator)是自主成分，Ds (dissociation)是非自主成分。在 Ac 成分的激活下，Ds 成分可以转座到新的靶点上，使靶点附近的基因失活或改变表达活性。Ds 成分的插入也能造成染色体断裂。不同的 Ds 成员有不同的大小，可以从 0.4Kb 到 4 Kb；而 Ac 全长 4.5 Kb。该系统两端均为 11 bp 的完美倒转序列；而靶点的同向重复则可以 6～8bp。Ac 成分上的两组 ORF 从中间开始向着两端伸展，这很象细菌转座元 Tn 3。Ac-Ds 成分的结构示于图 11-36。

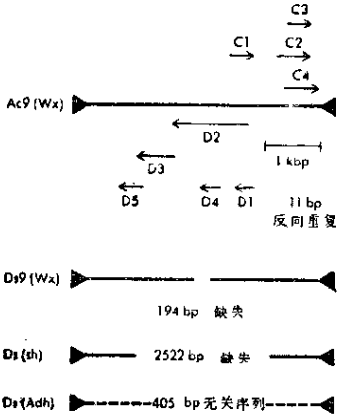



图 11-36 Ac 成分和若干 Ds 成分的结构。其中两种 Ds 成分为 Ac 成分的缺失突变产物。还有一种是由 405bp无关序列加上 Ac 和 Ds 所共同具有的 11 bp 末端倒转重复所组成，它插入乙醇脱氢酶基因中并使之失活。$C_1$-$C_4$ 和 $D_1$-$D_5$ 表示 Ac 成分上的两组开放阅读框。

鉴于 Ac 成分和 Ds 成分的具体转座机制还不清楚，但它们的转座常常伴随着在给点位置上的消失，这与细菌转座元的转座机制不同。当它们在给点位置上切除以后，留下 6～8bp 的同向重复序列(有时不完全相同，可能是切除前或切除后产生的突变而引起的)。这就是 Ac 或 Ds 成分曾经到过这里而留下的脚印。

# 第八节　真核生物中的转座成分(二)——还原毒病[46~61]

## 一、通过 DNA 中间体复制其基因组 RNA

还原病毒与其寄主细胞有着极为密切的关系。尽管还原病毒基因组 RNA 为正链RNA，但是无论是其基因产物的表达还是后代病毒的产生都必须通过一个 DNA 中间体。这个DNA中间体以整合状态存在于细胞基因组中。这个中间体叫做原病毒(provirus)。原病毒作为基

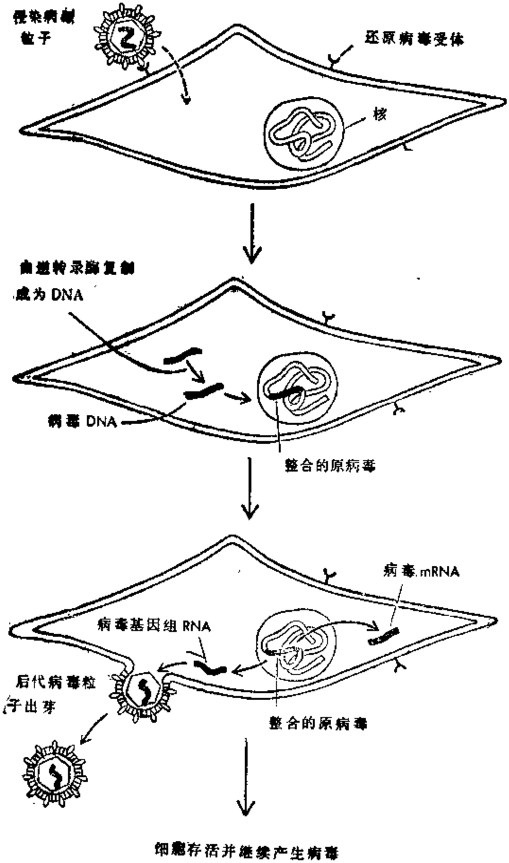



图 11-37　还原病毒生活史略图。侵染的病毒粒子以其糖蛋白与细胞上的特定受体结合，然后进入细胞质，在细胞质中进行反转录和双链 DNA 的合成，然后进入细胞核并整合到基因组中成为原病毒。由原病毒转录产生 mRNA 和后代噬菌体的基因组 RNA。

因组的一部分，可以随着细胞基因组的复制和细胞分裂而传递下去。新的病毒基因组RNA或其 mRNA 均由未经修饰的寄主转录机器（RNA 聚合酶Ⅱ）以原病毒 DNA 为模板转录产生。病毒蛋白质将病毒 RNA 包装成病毒核粒（nucleocapsid）以出芽方式“分泌”到细胞外；在出芽过程中，获得了由细胞膜和病毒糖蛋白构成的外膜（envelope）而成病毒粒子（virion）。在通常情况下，寄主细胞并不被杀死。一个含有原病毒的细胞不能再被同种病毒重复感染，这是因为原病毒所产生的病毒外膜糖蛋白已占据了细胞的相应受体。上述过程示于图 11-37。

需要指出，侵染的还原病毒 RNA 在细胞质内进行的反转录和双链 DNA 的合成以及整合过程全由逆转录酶所催化，而病毒 RNA 基因在刚侵入细胞后又不能表达，因此逆转录酶分子必定存在于病毒粒子中，在侵染时与基因组 RNA 一起进入细胞质内。这是所有的还原病毒的共同特征。

## 二、还原病毒的基因组结构

在还原病毒粒子的外膜内包裹着20面体的核粒，核粒中存在着两个拷贝的病毒单链RNA以及与之部分互补的寄主的两个 tRNA 分子。由于每个病毒粒子中有两个单链 RNA 拷贝，因此其遗传行为类似于二倍体。一个拷贝的某个基因失活以后，另一拷贝可以起反式互补作用。然而，它们存在两个拷贝的实际生理意义还不清楚，可能是有利于损伤修复和增强重组效率。

还原病毒基因组 RNA 长 7 000～10 000 碱基，现在已知的十几种还原病毒的基因组结构均很相似。典型的还原病毒有三个蛋白质编码基因：*gag*（group-specific antigen）编码一个多蛋白质前体，然后经翻译后处理成为几种核粒蛋白质；*pol* 与 *gag* 共同编码一个多蛋白质，酶切以后产生逆转录酶，整合酶以及蛋白酶；*env* 编码外膜糖蛋白的前体。最近还发现人类 T 淋巴细胞瘤病毒 HTLV Ⅰ 和 HTLV Ⅱ 以及 AIDS 病毒的基因组中含有第四类编码转录激活因子的基因（*px*，*lor*，*tat* 等），这些转录激活因子能极大地增加病毒基因的表达。AIDS 病毒还含有另一些在其他还原病毒中所没有的基因。此外，在一些还原病毒基因组中还包含能迅速引起肿瘤的癌基因（*onc* 基因），如 RSV 中的 *src* 基因。

还原病毒基因组两端还有很长的非编码序列。在 5′ 端有帽子结构，在 3′ 端有 poly(A) 尾巴。除了这两个转录后加工的产物之外，基因组本身在两端有完全相同的 10～80 碱基正向重复序列 R。在两个 R 片段的内侧，5′端部分为 80～100 碱基长的 $U_5$(unique to the 5′ end)，3′ 端部分为 170～1 250 碱基长的 $U_3$(unique to the 3′ end)。在 $U_5$ 的内侧为引物结合位点 PB(primer binding)，这个引物就是在 RNA 基因组复制时从寄主细胞中偷来的 tRNA。在不同的还原病毒中 tRNA 的种类也不同（如 RSV 中是 $tRNA^{Trp}$），然而这些 tRNA 都是通过 3′ 末端部分与 PB 位点互补以便在侵入细胞质后引发 DNA 的合成。逆转录酶和所有的DNA 聚合酶一样，都不能从头起始一条新 DNA 链的合成，而必须在引物的 3′-OH 上进行延伸。图 11-38 表示几种有代表性的还原病毒的基因组结构。

## 三、还原病毒基因组 RNA 到原病毒 DNA 的转变过程

还原病毒粒子中的逆转录酶在进入细胞以后面临着一项极其艰巨的任务，那就是将单链 RNA 分子转变为双链 DNA 分子并整合到寄主细胞基因组中以使 RNA 聚合酶 Ⅱ 能够进行转录。这个整合状态的原病毒并非是 RNA 基因组的精确拷贝，而是在 5′ 端增加了 $U_3$，在 3′

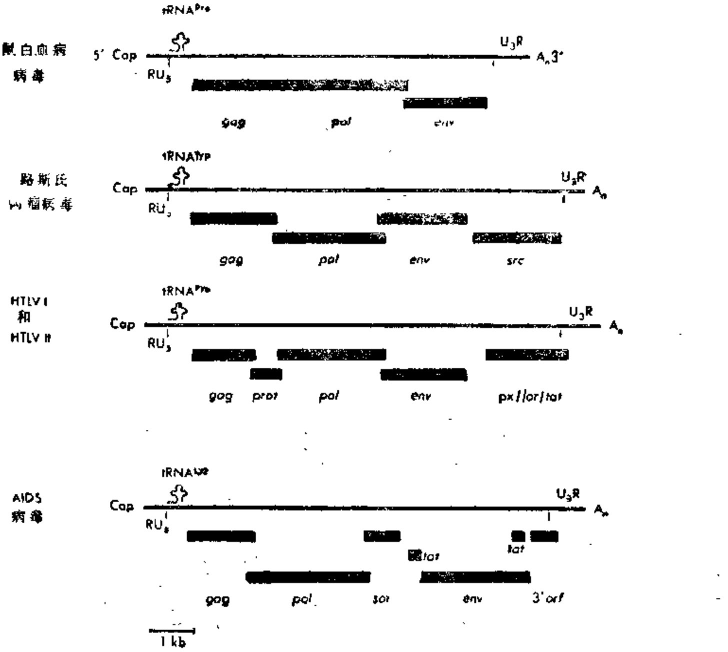
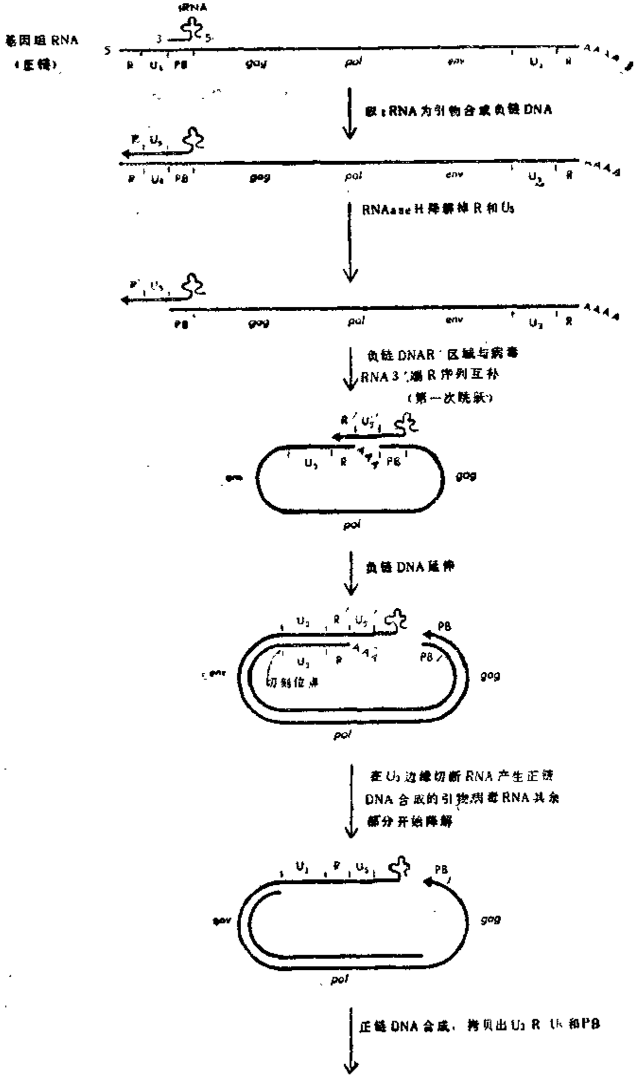



图 11-38 四种还原病毒基因组的结构。大多数还原病毒都像鼠白血病病毒(MLV)那样仅有 *gag*、*pol*、*env* 三个基因构成；少数含有 *onc* 基因，如 RSV；而 HTLVI，HTLVⅡ和 AIDS 病毒有更为复杂的结构。注意 AIDS 病毒的 tat 基因中含有一个内元。

端增加了 $U_5$，产生了两个完全相同的正向重复序列 $U_3RU_5$，称为长末端重复 LTR(long terminal repeats)。在长末端重复中有着转录起始信号和 3′ 端切断并加 poly(A)尾巴的信号。

病毒粒子进入细胞质以后，基因组 RNA 则释放出来，同时逆转录酶也释放出来。逆转录酶以 tRNA 为引物，以 $U_5R$ 为模板，反向转录产生 cDNA (大约 100～200 核苷酸)。在逆转录酶到达基因组 5′ 末端时，该酶的 RNAase H 活性切除了刚才拷贝过的模板。于是新生的 DNA 序列 $U_5'R'$ 则借助于 R′ 与 R 的互补关系与 3′ 末端的 R 序列相结合 (第一次跳跃)，因此产生了新的模板 ——引物搭配。利用这新的模板——引物搭配，逆转录酶能产生全部 RNA 基因组的互补链。为了转变成双链 DNA，在 RNA 模板 $U_3$ 序列的左边有一个切割位点，从这里切断 RNA，并以其 3′OH 为引物，开始了正链 DNA 的合成，产生了正链 DNA 的$U_3RU_5PB$部分。然后出现了第二次跳跃，即 tRNA 被 RNAase H 活性所分解，正链上的 PB 序列于是与负链上的 PB′ 序列互补，然后 RNAase H 活性又降解了整个基因组 RNA。最后将两条 DNA 链的合成进行到底，则产生具有长末端重复 $U_3RU_5$ 的双链线形 DNA。这一过程示于图11-39。全部过程都由逆转录酶所催化。在这个过程中该酶必须有四种活性：RNA 所指导的 DNA 聚合酶活性，DNA 所指导的 DNA 聚合酶活性，RNAase H 活性和位点特异性的 RNA 切刻酶活性(位于 $U_3$ 的 5′ 端)。


基因组RNA
(正链)
tRNA
取tRNA为引物合成负链DNA
RNAase H降解掉R和U5
负链DNAR'区域与病毒
RNA 3'端R序列互补
(第一次跳跃)
负链DNA延伸
切割位点
在U3边缘切断RNA产生正链
DNA合成的引物,病毒RNA其余
部分开始降解
正链DNA合成，拷贝出U3 R U5和PB
gag
pol
env
PB

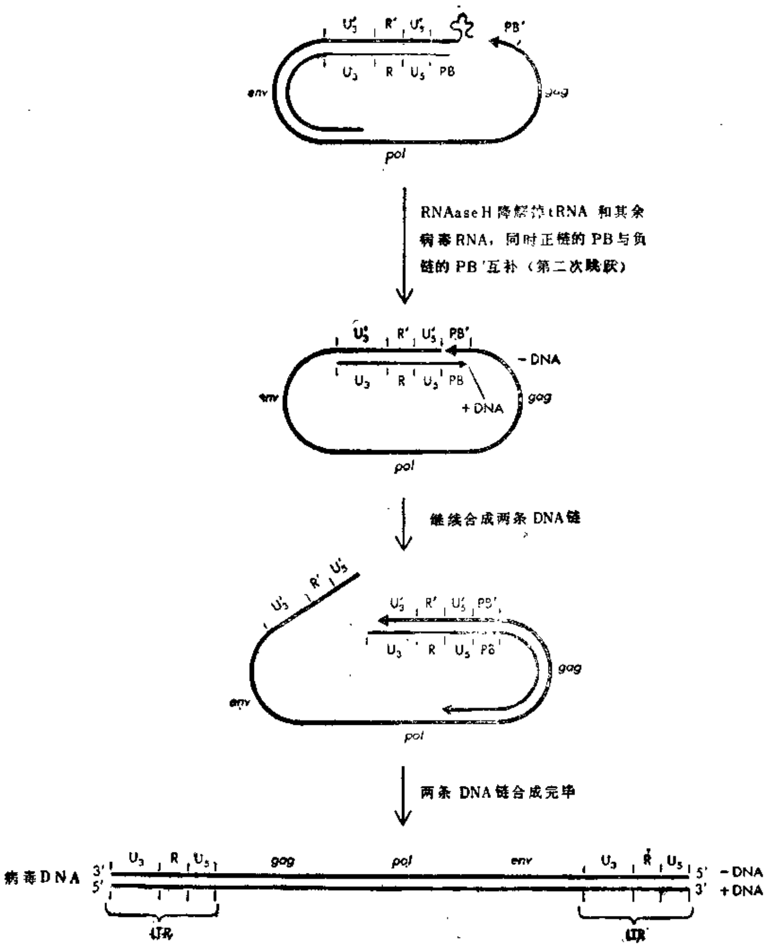



图 11-39　还原病毒的线形双链 DNA 的合成过程。这一过程所需的各种酶活性全部由逆转录酶所提供。

在细胞质中合成线状的原病毒 DNA 以后，接着就进入细胞核，在细胞核内转变为环形分子。有的环形分子含有一个 LTR，可能是通过线状双链 DNA 的两个 LTR 之间的同源重组后产生；有的环状 DNA 分子含有紧密相连的同向 LTR，是通过线形双链 DNA 末端直接连接而产生的，这种连接酶可能是寄主细胞核内的成分。

还原病毒的整合过程还不十分清楚。这一过程需要病毒编码的整合酶。整合到寄主细胞基因组中的原病毒 DNA 总是比线形原病毒 DNA 每端少两个 bp，其余则完全相同。对于靶

点的选择是随机的。靶点序列产生 4～6 bp 的倍生。每个被侵染的细胞通常有 1～10 个原病毒整合在细胞基因组中。所有原病毒 DNA 上与细胞基因组 DNA 相连的两端序列都是 5′TG……CA3′。而且，酵母的 Ty 成分，果蝇的 *copia* 成分以及 Mu 噬菌体的两端也都是同样的序列，可见它们之间是有进化上的密切关系。还原病毒的整合过程示于图 11-40。

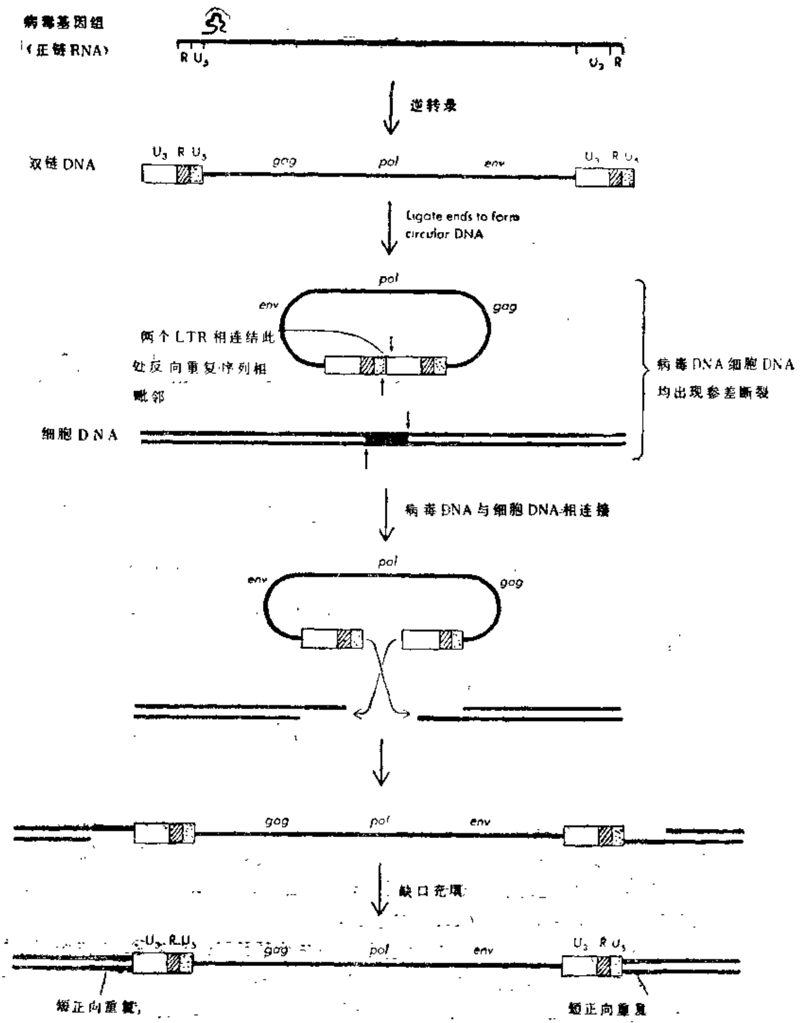



图 11-40 还原病毒与寄主细胞 DNA 的整合过程模型

## 四、原病毒的基因表达

LTR 不但提供了整合所必须的末端，而且提供了转录和转录后处理的信号。这些信号是非常有效的，加上某些还原病毒有自己编码的转录激活因子，有时在被侵染的细胞中，总 mRNA 的 10% 竟是病毒 RNA；LTR 的 $U_3$ 序列中含很强的启动子，在 −25 和 −80 具有典型 TATA 序列和 CCAAT 序列。帽子位点位于 $U_3$ 和 R 交界处。在远上游有转录的增强子，常常以两个同向重复的长 70～100 bp 的序列存在。还原病毒的增强子与 SV 40 的增强子具有很高的同源性，有的甚至可以代替 SV40 的增强子(图 11-41)。

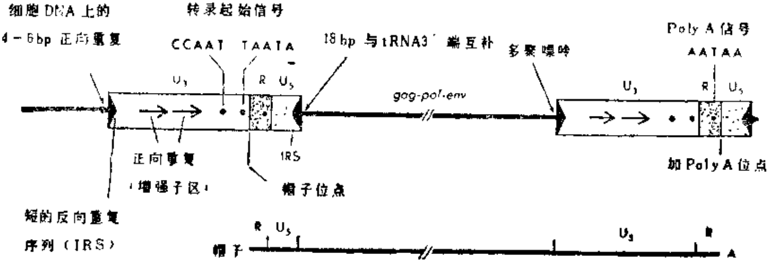



图 11-41 还原病毒 LTR具有原病毒基因表达全部信号。本图所示为鼠白血病病毒(MLV)。所有的还原病毒以及酵母和果蝇中的还原转座元均具有这些共同的特征。

原病毒基因由 RNA 聚合酶 Ⅱ 所转录，转录后处理过程也全是由寄主的有关机构进行。转录产物在 R 序列的 3′ 端切断并加 poly(A)尾巴，它们有两种不同的命运：一种是转运到细胞质中作后代病毒的基因组 RNA，另一种是作 mRNA 而翻译产生病毒所必须的蛋白质。还原病毒的基因表达有着复杂的转录后处理和翻译后处理，这里不能详细叙述，只将其主要方面示于图 11-42。

## 五、还原病毒的癌基因及其起源

绝大多数还原病毒可以在寄主体内诱发肿瘤的产生。同时许多还原病毒还能将组织培养的正常动物细胞诱导成生长不受限制的状态，这一过程叫做转化(transformation)。只含有 *gag*、*pol* 和 *env* 基因的还原病毒仅能以很慢的速度在体内诱发肿瘤，不能转化培养细胞，这些病毒是弱致癌(weakly oncogenic)还原病毒。有些还原病毒能够转化组织培养的细胞，并能迅速地在动物体内诱发肿瘤，它们叫做强致癌(highly oncogenic)还原病毒。除了 *gag*、*pol* 和 *env* 外，强致癌病毒基因组还含有另一个基因，即癌基因(oncogene，通称 onc)。比如研究较多的强致癌病毒——鸡劳斯肉瘤病毒的 *src* 基因。

有些强致癌还原病毒中 *gag*、*pol* 或 *env* 基因由缺失等原因而造成失活。这种缺陷型病毒本身难以完成侵染周期，必须与同种的正常病毒(称为辅助病毒，即 helper virus)共侵染才能复制繁殖自己。所有的强致癌还原病毒中，除了 RSV 以外，都属于缺陷型病毒。

从还原病毒的 *onc* 基因产物分布来看，大多数都分布在细胞质膜附近，可能是主要通过影

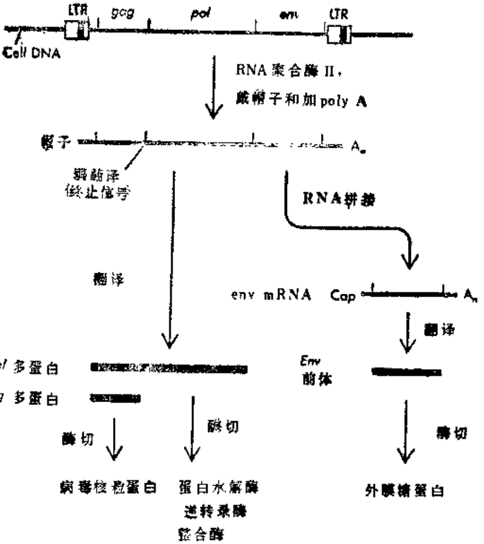



图 11-12 还原病毒基因的转录和翻译。只有一小部分原初转录物被拼接产生外膜蛋白的 mRNA。在其余的作为 mRNA 使用的原初转录物中，有 90% 的翻译产物在 *gag* 基因末尾的终止密码处终止，还有 10% 可以通过无义抑制产生通读，产生 gag-pol 多蛋白质，再通过翻译后加工产生逆转录酶，整合酶和蛋白酶。

响细胞对外界调控信号的反应能力和方式而引起细胞转化。蛋白质激酶，特别是酪氨酸特异性的蛋白质激酶(即在酪氨酸羟基上进行磷酸化)，生长因子和激素的受体，生长因子等类 *onc* 基因产物据绝大多数。根据基因产物的分布和功能，*onc* 基因可分为如表 11-7 所示的六类。

显然，强致癌病毒是由弱致癌病毒获得了 *onc* 基因后演化而来的，那么 *onc* 基因又来自何处呢？以 v-*src* 基因为探针和正常动物细胞中的 DNA 作分子杂交，结果发现鸡正常细胞DNA中的确存在着可与 v-*src* 基因杂交的基因(称为 c-*src*，c 意指细胞)。分子杂交试验还表明哺乳类、鱼类甚至果蝇的 DNA 中也有 c-*src* 基因。进一步分析证明，除了 c-*src* 基因有内元而v-*src* 基因没有外，二者在碱基序列上十分相似。在另一个实验中，用失去一小段 v-*src* 基因的 RSV 侵染鸡细胞，所产生的后代病毒中一部分重新获得了完整的v-*src*基因。显然，v-*src* 的缺失部分是由细胞的c-*src*基因提供的。以上结果说明还原病毒的 v-*onc* 基因来自寄主细胞的 c-*onc* 基因。除了 RSV 外，在许多鸟类和哺乳动物的强致癌还原病毒中已分别发现了 20 余种与正常的细胞核基因密切相关的 v-*onc* 基因，它们的名称来自最早被发现时所在病毒的名称，如 *src* 来自 Rous Sarcoma virus，*fes* 来自 feline sarcoma virus (猫肉瘤病毒)。目前许多人类癌细胞的 *onc* 基因已经清楚。例如膀胱癌基因长 5.4 Kb，结肠癌及肺癌基因长 45 Kb。尽管它们的长度差异很大，但却具有类似的内元——外元结构，其蛋白质产物均为 21 000 道尔顿，在序列上有一定的相似之处，这些蛋白质位于癌细胞的细胞膜外侧，与强致癌还原病毒的 *onc* 基因产物密切同源。人类膀胱癌基因与Harvey肉瘤病毒的*ras*基因相似，肺癌基因与 Kir-

表 11-7 *onc* 基 因 家 族

| *onc* 基 因 | 蛋白质的亚细胞分布 | 性 质 或 正 常 功 能 |
|---|---|---|
| Ⅰ. 蛋白质激酶: | | |
| *src* | 细胞质膜 | 酪氨酸专一性的蛋白质激酶 |
| *yes* | 细胞质膜 | 酪氨酸专一性的蛋白质激酶 |
| *fgr* | ? | 酪氨酸专一性的蛋白质激酶 |
| *abl* | 细胞质膜 | 酪氨酸专一性的蛋白质激酶 |
| *fps*(*fes*) | 细胞质 | 酪氨酸专一性的蛋白质激酶 |
| *erbB* | 细胞质膜(跨膜) | 酪氨酸专一性的蛋白质激酶/EGF 受体 |
| *fms* | 细胞质膜(跨膜) | 酪氨酸专一性的蛋白质激酶/CSF-1 受体 |
| *ros* | 细胞质膜(跨膜) | 酪氨酸专一性的蛋白质激酶 |
| *kit* | 细胞质膜 | |
| *mos* | 细胞质 | 丝氨酸/酪氨酸蛋白质激酶 |
| *raf*(*mil*) | ? | 丝氨酸/酪氨酸蛋白质激酶 |
| Ⅱ. GTP结合蛋白: | | |
| *H-ras* | 细胞质膜 | 有 GTPase 活性的 GTP 结合蛋白 |
| *K-ras* | 细胞质膜 | 有 GTPase 活性的 GTP 结合蛋白 |
| Ⅲ. 生长因子: | | |
| *sis* | 分泌 | 由 PDGF 基因衍生 |
| Ⅳ. 核蛋白: | | |
| *myc* | 细胞核 | |
| *myb* | 细胞核 | |
| *fos* | 细胞核 | |
| *ski* | 细胞核 | |
| Ⅴ. 激素受体: | | |
| *erbA* | 细胞质 | 甲状腺激素受体 |
| Ⅵ. 转录因子: | | |
| *jun* | 细胞核 | RNA 聚合酶 Ⅱ 的转录激活因子AP1, 结合位点为 TGACTCA |

sten 肉瘤病毒的 *ras* 基因相似。这些人类 *onc* 基因由正常细胞的原癌基因(protooncogene)突变而来。有时突变只涉及到一对碱基的代换。例如人类膀胱癌原癌基因中的甘氨酸密码子变为癌基因中的缬氨酸密码子(GGC→GTC)。

既然正常细胞中含有 c-*onc* 基因，那么为什么它们极少诱发癌肿而病毒的 v-*onc* 却有强烈的致癌作用呢？一种假说认为，c-*onc* 基因编码关键调节蛋白，正常状态下 c-*onc* 的表达活性很低，其产物的量足以维持正常生长但不足以破坏细胞的调控平衡，而逆原病毒的 LTR 具有很强的启动子活性，由它控制的 v-*onc* 基因表达的效率高，产生过量的产物，破坏了调控平衡，使细胞生长不再受到限制，也就是说 c-*onc* 和 v-*onc* 功能上的不同在于它们表达活性的不同。另一种假说认为，v-*onc* 和 c-*onc* 基因的碱基序列有一些小的差异，这种差异决定了 v-*onc* 的强烈致癌作用，与之相反，c-*onc* 的主要功能是调节细胞生长。支持前一种假说的证据来自对弱致癌病毒机理的研究。前面讲过，不含 v-*onc* 基因的弱致癌逆原病毒能够缓慢地诱发肿瘤，在这些肿瘤细胞中，原病毒几乎总是插入在 c-*onc* 基因的上游贴近处，由于原病毒可以提供强启动子功能或增强子功能，因此附近的 c-*onc* 基因的表达活性可以大大增加，这在效果上等同于处于原病毒内的 v-*onc* 的功能。这里发生变化的不是 *onc* 基因本身，而是它的表达活性。支持后一种假说的证据主要来自对癌细胞中和正常细胞中 *onc* 基因序列的比较。如前文

中提到过的人类癌细胞的 3 种 *onc* 基因均与正常细胞的 c-*onc* 基因有微小差异，但目前尚不知这些微小差异是否足以致癌。

综合以上讨论，可以认为至少相当一部分肿瘤的形成是由于：(1) 向细胞中导入处在强启动子控制下的 *onc* 基因(如 v-*onc*)，强致癌还原病毒的致癌原因可能在于此；(2) 向细胞中导入能够大大提高 c-*onc* 基因表达活性的启动子或增强子，弱致癌还原病毒的作用可能在于此；(3) 产生提高 c-*onc* 基因表达活性或改变 c-*onc* 基因产物功能的突变，各种物理、化学诱变剂的作用可能在于此。

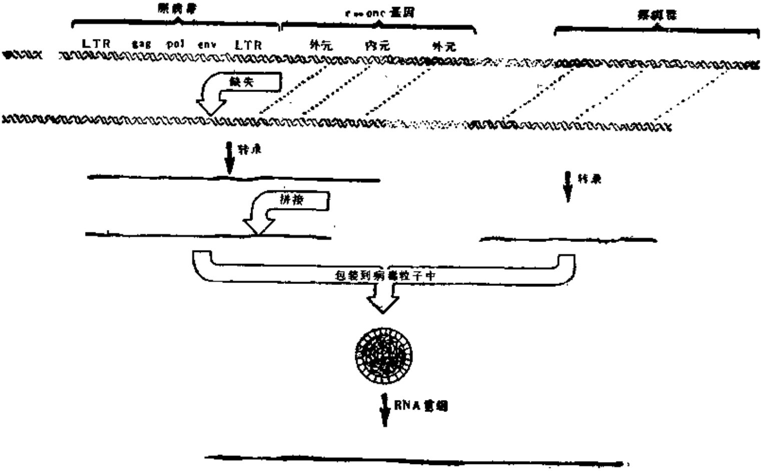



图 11-43 还原病毒中 v-*on* 基因起源的可能机制

还原病毒是如何从寄主细胞中获得 *onc* 基因的呢？目前对这一问题的解释见图 11-43。还原病毒整合到寄主染色体的 c-*onc* 基因上游；然后染色体上出现一段缺失，使原病毒和 c-*onc* 融合到一起，这样 c-*onc* 基因的转录便处在原病毒左端 LTR 强启动子的控制之下，所合成的 RNA 分子包括病毒序列和 c-*onc* 序列两部分，经过 RNA 拼接去掉内元。在细胞内含有正常原病毒拷贝的情况下，这种 RNA 分子可以和正常病毒 RNA 分子一起包装进病毒颗粒，在下一轮侵染中，上述病毒中的两种 RNA 分子都被反向转录成线状双链 DNA，再通过正常病毒 DNA 和含有 *onc* 基因的病毒 DNA 之间的重组(很可能是非同源的异常重组)，在 *onc* 基因的右侧加上一个 LTR 重复单位，这样的病毒 DNA 便可以整合到寄主 DNA 中并繁衍自己。在此之后，病毒中的 *onc* 基因和细胞中的 *onc* 基因各自积累突变而分道扬镳，最后分别演化成今天的 v-*onc* 基因和 c-*onc*基因，当然，还原病毒同样能够以上述方式捕获细胞中的其它基因，但这往往会给病毒的生存带来不利，因此通常很快就被负向选择从群体中淘汰掉；而 *onc* 基因能促进细胞生长，加快分裂，具有选择优势，所以含有 v-*onc* 基因的还原病毒能一直存在繁衍。

### 六、还原病毒和转座元的关系

还原病毒和转座元的相似性引起了人们的广泛注意。有一部分人认为，还原病毒由古代的

转座元衍生而来。由于现在所有的还原病毒都具有相似的结构，因此它们很可能来自同一个祖先转座元。可以设想某种原始的类似插入序列的转座成分插入到寄主的某种原始的DNA聚合酶基因(*pol*)两侧，由此产生的转座成分具有LTR-*pol*-LTR的结构，其中的*pol*基因在进化中逐渐演化成逆转录酶基因，这种转座成分再通过转座作用从寄主DNA或其它来源的DNA中捕获负责合成病毒核粒和被膜蛋白质的基因，即今日*gag*和*env*基因的前身。此时，含有*pol*、*gag*和*env*基因的转座成分便具备了作为病毒的必要条件，形成了以还原病毒的侵染繁殖方式以扩增自已的能力。在由转座元到还原病毒的进化中，原先的转座能力也许已随着病毒侵染能力的出现而逐步消失，但也许目前依然起着作用。现在尚无充足证据证明原病毒的整合作用是病毒线状双链DNA分子本身的直接插入还是像典型的转座元那样伴随着病毒DNA分子的复制；也无证据证明原病毒是否能像典型转座元那样以转座的方式在寄主DNA上插入新的拷贝。不管怎样，由转座元向病毒的进化使有关的DNA序列获得了在细胞间，个体间和群体间转移的有效手段，大大提高了这些DNA序列拷贝扩增的广度和效率，因此在进化上是有利的。

还有人提出，果蝇的*copia*成分可能是目前尚未发现的还原病毒。在组织培养的果蝇细胞中已经发现了类似还原病毒的颗粒，其中含有5 Kb长的RNA，和*copia*等长。这些类病毒颗粒中含有逆转录酶和长为5 kb的RNA分子，更重要的是类病毒颗粒中的5 Kb RNA能够与*copia* DNA杂交。这些发现强有力地支持了*copia*成分是某种还原病毒的原病毒这一观点。联系到前文中关于还原病毒由古代转座元衍生而来的讨论，可以认为*copia*成分可能是正在向还原病毒演化的转座元，它已具备了编码逆转录酶及病毒颗粒蛋白质的基因，但还不完备，目前尚不能形成具有侵染能力的还原病毒颗粒，而只能在细胞内采取还原转座的方式(以RNA为媒介)进行繁衍。另一部分人认为，*copia*成分是失去了形成病毒颗粒能力的古代还原病毒的衍生物。如果是这样，则表明转座元和还原病毒(可能还有其他病毒)之间可以互相转变。

## 第九节 真核生物中的转座成分(三)——非病毒返座元[36,62]

### 一、概 述

真核生物中的非病毒返座元均起源于RNA聚合酶Ⅱ和RNA聚合酶Ⅲ的转录产物，而且绝大部分是经过完全的转录后处理的RNA产物。迄今所发现的例外只有两个：鼠类的前胰岛素原(preproinsnlin)Ⅰ基因是半处理的mRNA的返座产物，而且是有转录活性的返座基因；另一个是不完全处理的U2 snRNA假基因。通过返座作用产生的假基因称为返座假基因(retropseudogene)。来自于完全处理的mRNA返座假基因全部具有poly (A)尾巴并缺少内元，成熟mRNA的碱基序列很少受到触动，但有时某一端(5′端或3′端)有被删节的现象。非病毒返座元在返座时产生靶点序列的倍生，长度一般介于7-21bp；然而，它们本身的正向重复序列或长或短，甚至完全没有。有趣的是，靶点序列倍生的长度对于返座成分来说并不是特征性的，例如人类Alu序列两端的靶点倍生长度可以从9bp一直到21bp。

非病毒返座元的最大特点乃是它们没有共同的结构特征。其长度可以从33 bp(U 2.1假基因)一直到6 Kb以上(LINE1)。尽管返座元常常有3′末端poly (A)序列，但是，大多数snRNA和偶蹄类的SINE重复序列都没有poly(A)序列。显然，由于RNA聚合酶Ⅱ和Ⅲ的产物均可以返座，可见5′端结构(如帽子或三磷酸基团)和3′端结构[如poly(A)或寡聚U]

都不是它们返座作用所必须的前提条件。非病毒的返座元可能都是被动返座，因为它们都缺乏与返座作用有关的基因。

## 二、RNA 聚合酶Ⅱ转录产物的返座假基因和返座基因

如上所述，RNA聚合酶Ⅱ转录产物的返座假基因几乎全部是来自于完全处理的mRNA，3′末端总是有 poly(A)序列，缺乏任何内元，通常延伸至5′端的帽子位点。但5′端删节的现象也并不少见。可能的原因有错误的拼接，或返座以后又有别的转座成分的插入。当然也不能排除逆转录不彻底的可能性，但是大量的 RNA 聚合酶Ⅱ转录产物的返座假基因都是全长的，因此不像是逆转录不彻底造成的。

由于 RNA 聚合酶Ⅱ的启动子总是在转录序列的上游，因此，几乎全部的这类返座产物都是返座假基因。由于假基因不受选择压力的制约，逐渐积累了很多突变。然而也不能排除这样的可能性，返座元返座于某一个外源启动子的下游而获得转录活性，鸡的一个缺乏内元的钙调蛋白基因可能就是这种情况，不过，还没有确凿证据。而大鼠和小鼠前胰岛素原I基因是由其祖先基因启动子上游开始的某种连续转录物经部分转录后处理(加 poly A 尾巴和除去一个内元)然后返座而产生的。将鼠类的前胰岛素原Ⅰ和Ⅱ的基因与鸡、狗、豚鼠，人类的单拷贝(在单倍体基因组中)前胰岛素原基因进行比较，就可以知道鼠类的Ⅱ基因是原型基因，因为它与其他动物的单拷贝基因一样，具有两个内元，小的一个位于5′端非编码序列中，大的一个位于C肽编码序列中。而Ⅰ基因中只含有5′端非编码序列中的小内元。Ⅰ基因和Ⅱ基因都具有较高的转录活性。这两个基本相同的基因一定受到选择压力的制约，因为二者编码的A链和B链完全相同，而C肽序列已经积累了突变(C肽通过翻译后处理而被除去)。鼠类前胰岛素原Ⅰ基因可能并非是真核生物唯一的有功能的返座基因，摇蚊(*Chironomus*)的无内元珠蛋白基因可能又是一例。

奇怪的是，在各种真核生物中，哺乳动物的返座现象最为常见，例如哺乳动物有20～30个微管蛋白返座假基因，而果蝇和鸡则没有。更为显著的例子是，哺乳动物中3-磷酸甘油醛脱氢酶的返座假基因十分丰富，在人中有25个拷贝，而在大鼠和小鼠中则在200个拷贝以上。

在 RNA 聚合酶Ⅱ转录物的返座元中，有一类拷贝数很多、序列很长的返座元，叫做长散布重复序列(long interspersed repeated sequences)，简称 LINEs，或称为 LINE 1 家族(以 LINE 1 为代表)。在哺乳动物中，LINE 1 家族有20 000～50 000个拷贝，长6～7 Kb，3′末端有富含A的序列，前面有时还有多聚腺苷酸化的信号。也有人认为 LINE 1 家族是位于该成分上游的 RNA 聚合酶Ⅲ启动子所进行转录的通读产物返座而成的。LINE 1 家族通常有一个或几个开放阅读框，其产物还不清楚。有人推测，LINE 1 家族可能并没有被“冻结”，可能会继续进行返座作用。

## 三、RNA 聚合酶Ⅲ转录产物的返座元

许多 RNA 聚合酶Ⅲ的转录产物都可以返座。这些返座元一般都较短，大约70～300bp，但是拷贝数极高，通常在100 000以上。这些返座元称为短散布重复序列(short interspersed repeated sequences)，简称 SINEs，一个 SINE 家族中的各成员之间，都有相似或相同的5′末端和3′末端。SINEs 通常使靶点出现7～21 bp的倍生。大多数SINE家族都有一个有功能的内部启动子(RNA 聚合酶Ⅲ的启动子)。这种内部启动子能够转录相邻的染色体序列。

人类的 Alu 序列家族和鼠类的 B1 序列家族可以作为 SINEs 的代表。

Alu 成分长约 300 bp，在人类基因组中约有 500 000 个拷贝，占人类 DNA 基因组的 5～6%。Alu序列家族，LINE 1 家族，THE 1 家族以及poly(CA)序列构成人类基因组中高度重复序列的绝大部分（由于Alu序列家族中又可分为有一定序列差异的数千个亚类，每一类有数百个拷贝，有人将Alu 序列归于中度重复序列，见第二章）。Alu序列为7SL RNA（见第七章）的返座产物，但与 7SL RNA 基因的差别是：Alu 序列由两个不完全相同的单体组成，其右单体比 7 SL RNA 基因少 155 bp，左单体还缺失另外 31 bp（图 11-44）。由于 7 SL RNA 是由

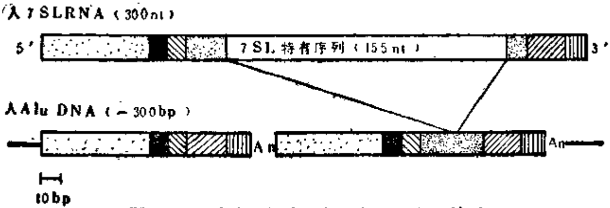



图 11-44 人类 Alu 序列与 7 SL RNA 的关系

RNA 聚合酶 Ⅲ 所转录的，因此在 Alu 序列中保留了两个 RNA 聚合酶 Ⅲ 的启动子。尽管 Alu 成分的右单体与 7 SL RNA 同源性更高，然而其中的启动子是没有活性的。只有在左单体中的启动子才有活性。其原因仍然是个谜。

奇怪的是，在人类基因组中尽管有 500 000 个 Alu 成分的拷贝，但是并未发现单体 Alu成分。而与 Alu 成分相当的鼠类 B1 序列家族却以单体存在。如此之多的 Alu 成分拷贝可能是来源于同一个祖先拷贝（也是由两个单体组成），然后通过返座作用形成新的 Alu 序列，这些新的 Alu 序列又可以转录而返座，因而产生了庞大的 Alu 序列家族。这一祖先拷贝中的两个单体可能来源于同一个单体的倍生，然后通过不同的缺失突变而演化产生的。今天的 Alu 成分是否还在扩散，有许多迹象表明，这种扩散过程仍然有可能在进行。

由于 RNA 聚合酶 Ⅲ 识别内部启动子，因此 Alu 序列的 5′ 端有固定的起点，这是不难理解的。在 Alu 序列扩散（返座）过程中，同样能够保持这样的 5′ 端序列。而 RNA 聚合酶 Ⅲ 的转录终止点通常处于一串 T 之中，因而 Alu 成分的转录产物的末端 U 能够与 Alu 成分本身所具有的 An 配对（在同一转录物中），从而引发互补 DNA 的合成。然后合成双链 DNA，再通过某种未知的中间物，最后整合到染色体的其他位置上（图 11-45）。

图 11-45 所示的返座机制仅系假说，还没有充足的证据，特别是自身引发的问题。尽管大多数RNA聚合酶 Ⅲ 的转录物末尾有寡聚 U，但是 RNA 聚合酶 Ⅱ 的转录产物经加工后末端只有poly(A)，这类转录物照样产生 cDNA 并返座。因此真核生物中的返座作用机制还有待于进一步的研究。还有，转录物自身引发的假说还无法解释为什么各种 Alu 序列中 An 不会缩短以至于消失的问题。一种可能性是最初合成的若干个碱基的 cDNA 会发生滑动，产生全长的与 An序列互补的 Tn，另一种可能性是在返座以后的任何时刻都可以通过某种机制使减少了的 An序列扩展。

鼠类的 B1 序列为单体 7SL RNA 返座产物，与 Alu 序列中的右单体很相似，只是除了缺失 155 个碱基对以外，还缺失另外 14 个 bp。真核生物基因组中还有许多小的重复序列，这些 SINE 大多是 tRNA 的返座产物。

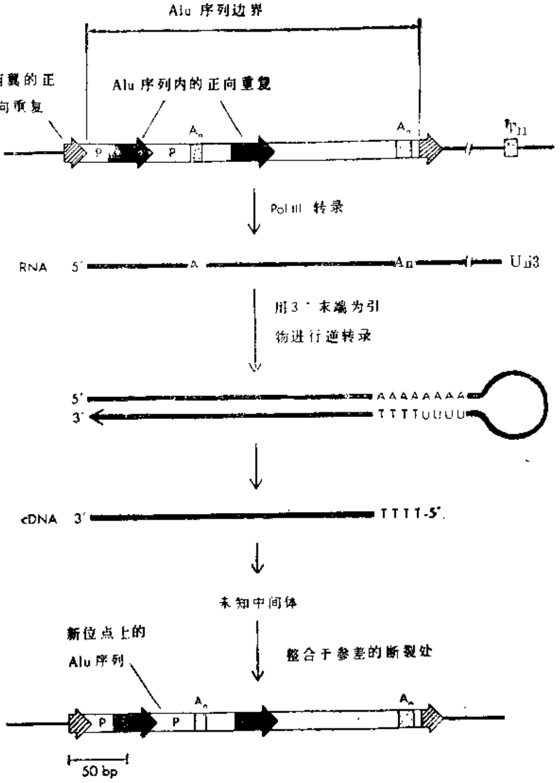



图 11-45 Alu 序列返座的可能机制

## 四、关于返座的机制

正如上面所述，Alu 序列的返座机制仅系假说。由于非病毒返座元都缺少编码逆转录酶和整合酶的基因，因此人们普遍认为，这类返座元的返座是被动的，而且可能是由一种共同的返座酶系统所催化，甚至 cDNA 合成的引发也具有共同的机制。在返座过程中，RNA 中间物是必不可少的。这就是目前人们公认的结论。需要进一步澄清的问题则多得多：

1. 在插入基因组之前 RNA 拷贝成 cDNA 还是 RNA 直接插入基因组 DNA 中然后拷贝出 cDNA？

2. 是什么引发 cDNA 的合成？除了自身引发之外，由第二个 RNA 分子或 DNA 也是有可能的。特别是，如果逆转录是在原位发生的，即 RNA 还没有脱离 DNA 模板时发生逆转录，则染色体 DNA 作为引物是很有可能的。

3. 逆转录酶活性的来源是什么？这是目前无法回答的问题。过去有人将这种酶活性归于还原病毒，但无法解释为什么鸡的返座成分要比哺乳动物少得多(鸡的 RSV 具有很强的逆转录酶)。还有，尽管果蝇和酵母都有内源性病毒返座成分并编码逆转录酶，但果蝇中返座假基因十分稀少，而酵母中显然不存在返座假基因。合理的推断是这些生物中的逆转录酶不足以产生返座，而必须是在适当的发育阶段，在适当的细胞的适当部位存在着适当的逆转录酶才能促成真核生物中的非病毒返座作用。联系到个体的发育，首先必须是配子的生成。精子发生过程，哺乳动物与鸟类，两栖类以至果蝇，都差不多；而哺乳动物卵子生成过程却不同于其他三类动物，因为存在较长的灯刷染色体阶段。如果返座作用主要发生在卵子生成过程中，则 Y 染色体上的返座基因应该很少。这有待于作进一步的考察。

4. 为什么染色体断裂处的单链末端可以从 7 变化到 21 个碱基而正向重复可以从 0 一直到 41 bp? 或许这种断裂是随机地发生于富含 A/T 区域，正如 Alu 成分的两侧均富含 A/T序列。

总之，真核生物特别是哺乳动物中的返座现象，也像其他种类的遗传重组一样，使真核生物基因组具有很大的流动性，既可以使某些基因失活，又可以激活某些基因，甚至产生新的基因活性。这在真核生物基因组的进化过程中起了很大的作用。返座作用的分子机制还有待于进一步的探索。

## 提　　要

从广义上讲，任何造成基因型变化的基因交流过程都叫遗传重组。然而本章所述为狭义的遗传重组即 DNA 分子内断裂-复合这一基因交流过程。这个过程也称为交换。根据对 DNA 序列和蛋白质因子的要求，可以把重组分为同源重组，位点特异性重组，转座作用和异常重组四类。

原核生物的 DNA 结合蛋白较少，其同源重组难以直接观察。真核生物的同源重组发生在减数分裂的极早期，参与联合的双方实际上是两条已经复制的同源染色体。非姊妹染色子体之间以侧面紧靠在一起，形成联会丝复合体。目前有多种模型旨在解释双链 DNA 之间同源重组的可能机制。其中比较流行的一种是最初由 Holliday 提出，后经 Meselson 和 Radding修改的链转移模型。重组过程中所形成的 Holliday 中间体有两种拆分方式，一种可以观察到交互重组的发生，另一种则观察不到交互重组的发生。但是不管拆分以何种方式进行，在两条DNA分子上都留下一段异源双链区。子囊菌中发生的基因转换现象说明这两种拆分机会均等。两个环状 DNA 分子重组的结果则产生 8 字型中间体，它同样有两种拆分方式。若用单切点的内切酶对 8 字型中间体进行切割，则形成 $\chi$ 结构。

39 KDa 的 RecA 是同源重组过程中最重要的蛋白质，又称为重组酶。RecA 具有单链 DNA 结合活性，双链 DAN 结合活性，NTPase 活性和促进互补单链复性的能力。RecA能催化单链 DNA 和双链 DNA 之间的联会。单链同化后异源双链区可逐步扩大。由 RecA 所催化的分枝迁移是单方向的，速度很慢，可允许 1% 以上的错配碱基对的出现。它与 DNA 呼吸作用所控制的 D-嘌呤的分枝迁移完全不同。

在同源重组中另一个重要的蛋白质就是 RecBCD。RecBCD 是一个多功能的酶，它具有依赖于ATP的单链和双链外切酶的性质(曾称为外切核酸酶V)，线形双链DNA螺旋酶活性以及 Chi 序列特异性的单链内切酶活性。当细胞内 DNA 具有平头末端时 RecBCD 就结合上去，利

用其ATPase活性水解ATP，从一端将双链DNA解开并向另一端移动。由于解旋活性较高而再旋活性较低，因此形成所谓兔耳结构。当 RecBCD 所识别的 Chi 位点 (5′GCTGGTGG3′) 进入单链嘈噗区域时，RecBCD 的特异的单链内切酶活性便在 Chi 位点 3′ 方向 4～6 核苷酸处切断。这样就产生了一条包含Chi 位点在内的单链尾巴，RecA 结合上去，然后与同源序列进行链交换。人们还推测 RecBCD 在 Holliday 中间体的拆分中起重要作用。

酵母 MAT 序列的转换是一种依赖于同源重组的位点特异性的序列代换。它既不同于一般的同源重组，也不同于位点特异性重组，也不同于转座作用。

位点特异性重组的一个典型例子是λ噬菌体与寄主的整合与切除。整合过程由整合酶所催化，同时还需要整合寄主因子。这样的整合反应是不可逆的。在切除过程中，除了需要整合反应中的全部因子外，还需要切除酶以及寄主编码的逆转刺激因子。这样的切除反应也是不可逆的。重组位点 BOB′ 和 POP′ 中的两个 15bp 的 O 序列完全相同。整合酶的作用方式十分类似于拓扑异构酶 I，它可能直接参与断裂和复合过程。

自从 50 年代初 McClintock 发现玉米的控制成分以来，各种生物中均发现了可转移成分即转座元。转座过程一般都伴随着复制过程，因而转座过程使得转座元的拷贝数增加。转座作用并不依赖转座元和靶点之间以及给点和靶点之间任何的序列同源性。转座过程由转座酶所催化，还需要拆分酶才能完成转座过程。转座过程中共合体的形成是探索转座机制的基本线索。细菌中的转座成分可分为插入序列、复合转座元、TnA转座元家族和可转移噬菌体等。

真核生物中的转座成分分为两类。一种与细菌中的情况一样，是从 DNA 到 DNA 的转移过程，仍称为转座作用，被转座的 DNA 单元也称为转座元。另一种是 RNA 所介导的转座作用，称为返座作用，被转座的遗传信息单元称为返座元。真核生物的转座元有果蝇的 P 成分，FB 成分，玉米的 Ac-Ds 转座系统等。返座元的种类要复杂得多，根据其结构和组成，可以分为病毒大家族和非病毒大家族。病毒大家族包括全部的还原病毒以及与还原病毒的原病毒相似的称为还原转座元的转座成分，如酵母的 Ty 成分，果蝇的 *copia* 成分等。

还原病毒与其寄主细胞有着极为密切的关系。尽管其基因组 RNA 为正链 RNA，但是无论是其基因表达还是后代病毒 RNA 的产生，都必须通过一个 DNA 中间体。这个中间体整合于细胞基因组中，称为原病毒。新的病毒基因组 RNA 或其 mRNA 均由未经修饰的寄主转录机构以原病毒 DNA 为模板转录产生。病毒蛋白质将病毒 RNA 包装成病毒核粒以出芽方式分泌到细胞外；在出芽过程中获得了由细胞膜和病毒糖蛋白构成的外膜而成为病毒粒子。在通常情况下，寄主细胞并不被杀死。一个含有原病毒的细胞不再被同种病毒重复感染，因为原病毒所产生的病毒外膜糖蛋白已占据了细胞的相应受体。侵染的还原病毒RNA在细胞质内进行的逆转录和双链 DNA 的合成以及整合过程全由逆转录酶所催化。这种酶与病毒 RNA 一起包装在病毒粒子中并在侵染时一同进入细胞。还原病毒的基因组有 *gag*，*pol*，*env* 三个基因，有的还携带癌基因。病毒的癌基因来源于细胞的原癌基因。有人认为，还原病毒是由古代的转座元演化而成，而 *copia* 等还原转座元正处在演化过程中。也有人认为 *copia* 成分是失去了形成病毒粒子能力的古代还原病毒的原病毒。总之，还原病毒与转座元之间存在着密切的进化关系。

非病毒返座元均起源于 RNA 聚合酶 Ⅱ 和 Ⅲ 的转录产物，而且绝大部分是经过完全的转录后处理的 RNA 产物。迄今所发现的返座元只有两个例外：鼠类的前胰岛素原 I 基因是半处理的 mRNA 返座产物，而且是有转录活性的返座基因；另一个是不完全处理的 U2 snRNA

假基因。非病毒返座无没有共同的结构特征,而且都缺乏与返座作用有关的基因,因此其返座过程可能是被动返座。而且可能是一种共同的返座酶系统所催化。至于这种酶的来源,目前尚无结论。过去人们曾将这种酶活性归于逆原病毒,但无法解释具有很强逆转录酶活性的鸡的返座成分要比哺乳动物少得多等现象。返座作用也像其他种类的遗传重组一样,使真核生物的基因组具有很大的流动性,加速了基因组的进化过程。

## 参 考 文 献


1. Meselson, M., and Weigle, J. J. (1961) "Chromosome Breakage Accompanying Genetic Reconstruction in Bacteriophage." *Proc. Nat. Acad. Sci.* 47: 857～868.
2. Holliday, R. (1964) "A Mechanism for Gene Conversion in Fungi." *Genetic Res.* 5. 282～304.
3. Radding, C. (1982) "Homologons Pairing and Strand Exchange in Genetic Recombination." *Ann. Rev. Genetics* 16: 405～437.
4. Conley, E. C., and West, S. C. (1989) "Homologous Pairing and the Formation of Nascent Synaptic Intermediates between Regions of Duplex DNA by RecA Protein." *Cell* 56: 987～995.
5. Smith, G. R. (1987) "Mechanism and Control of Homologous Recombination in *Escherichia coli*." *Ann. Rev. Genetics* 21: 179～201.
6. West, S. C. et al. (1983) "Enzymatic Formation of Biparental Figure Eight Molecules from Plasmid DNA and Their Resolution in *E. coli*." *Cell* 32: 817～829.
7. Cox, M. M., and Lehman, I. R. (1987) "Enzymes of Genetic Recombination." *Ann. Rev Biochem.* 56: 229～262.
8. Amundsen, S. K. et al. (1986) "*rec D*: The Gene for an Essential third subunit of Exonuclease V" *Proc. Nat. Acad. Sci*" 83: 5558～5562.
9. Smith, G. R. (1988) "Homologous Recombination in Prokaryotes." *Microbiol. Rev.* 52: 1～28.
10. Cunningham, R. P. et al. (1980) "Homologous Pairing in Genetic Recombination: RecA Protein Makes Joint Molecules of Gapped Circular DNA and Closed Circular DNA." *Cell* 20: 223～,235.
11. Shaner, S. L., and Radding, C. M. (1987) "Translocation of *Escherichia coli* RecA Protein from a Single-stranded Tail to Contiguous Duplex DNA." *J. Biol. Chem*, 262: 9211～9219.
12. Kowalczykowski, S. C. et al. (1987) "Properties of Duplex DNA-Dependent ATPase activity of *Escherichia coli* RecA Protein and Its Role in Branch Migration." *Proc. Nat. Acad. Sci.* 84: 3127～3131.
13. West, S. C. et al. (1980) Recognition of Duplex DNA Containing Single-Stranded Rigions by RecA Protein." *Proc. Nat. Acad. Sci.* 77: 2569～2573.
14. West, S. C. et al. (1981) "Heteroduplex Formation by RecA Protein: Polarity of Strand, Exchanges." *Proc. Nat. Acad. Sci.* 78: 6149～6153.
15. Ponticelli, A. S. et al. (1985) "Chi-Dependent DNA Strand Cleavage by RecBC Enzyme." *Cell* 41: 145～151.
16. Taylor, A. F. et al. (1985) "RecBC Enzyme Nicking at Chi sites during DNA Unwinding: Location and Orientation Dependence of the Cutting." *Cell* 41: 153～163.
17. Smith, G. R. (1983) "General Recombination." in *Lambda* II, ed. R. W. Hendrix et al., pp. 175～209, Cold Spring Harbor Lab, N. Y.
18. Stahl, F. W. et al. (1985) "In Phage $\lambda$, *cos* Is a Recombinator in the Red Pathway." *J, Mol, Biol*, 181: 199～209.
19. Strathern, J. N. et al. (1980) "The Structure and Organization of the Transposable Mating Type Cassettes in *Saccharomyces*," *Proc. Nat. Acad. Sci.* 77: 2839～2843
20. Nasmyth, K. A. (1982) "Molecular Genetics of Yeast Mating Type." *Ann. Rev. Genetics* 16: 439～500.
21. Klar, A. J. S. et al. (1984) "The Involvement of Double-Strand Chromosomal Breaks for Mating-Type Switching in *Saccharomyces. Cerevisiae*" *Cold Spring Harbor Symp. Quant. Biol.* 49: 77～88.
22. Klar, A. J. S., and Stratherm, J. N. (1984) "Resolution of Recombination Intermediates Generated During Yeast Mating Type Switching." *Nature* 310: 744～747.
23. Nash, H. (1975) "Integrative Recombination of Bacteriophage $\lambda$ DNA *in vitro*." *Proc. Nat. Acad. Sci.* 72: 1072～1076.
24. Nash, H. (1981) "Integration and Excision of Bacteriophage $\lambda$: The Mechanism of Conservative site Specific Recombination." *Ann. Rev. Genetics* 15: 143～167.

25. Weisberg, R., and Landy, A. (1983) "Site-Specific Recombination in Phage Lambda." in *Lambda* II, ed. R. Hendrix et al., pp. 211～250, Cold Spring Harbor Lab, N. Y.
26. Thompson, J. F. et al. (1987) "Cellular Factors Couple Recombination with Growth Phase: Characterization of a New Component in the λ-site Specific Recombination Pathway." *Cell* **50**: 901～908.
27. Cohen, S., and Shapiro, J. (1980) "Transposable Genetic Elements." *Sci. Amer.* **242**: 40～50.
28. Grindley, N. D. F. (1985) "Transpositional Recombination in Prokaryotes." *Ann, Rev. Biochem.* **54**: 863～896.
29. Morisato, D. et al. (1983) "Tn10 Transposase Acts Preferentially on Nearby Transposon Ends *in vivo*." *Cell* **32**: 799～807.
30. Heffron, F. et al. (1979) "DNA Sequence Analysis of the Transposon Tn3: Three Genes and three Sites Involved in Transposition of Tn 3." *Cell* **18**: 1153～1163.
31. Roberts, D. et al. (1985) "IS10 Transposition Is Regulated by DNA Adenine Methylation." *Cell* **43**: 117～130
32. Harshey, R. M. (1984) "Transposition Without Duplication of Infecting Bacteriophage Mu DNA." *Nature* **311**: 580～581.
33. Craigie, R., and Mizuuchi, K. (1985) "Mechanism of Transposition of Bacteriophage Mu: Structure of a Transposition Intermediate." *Cell* **41**: 867～876.
34. Simon, M. et al (1980) "Phase Variation in the Evolution of a Controlling Element." *Science* **209**: 1370～1374.
35. McClintock, B. (1951) "Chromosome Organization and Gene Expression." *Cold Spring Harbor Symp. Quant. Biol.* **16**: 13～57.
36. Weiner, A. M. et al. (1986) "Nonviral Retroposons: Genes, Pseudogenes, and Transposable Elements Generated by the Reverse Flow of Genetic Information." *Ann. Rev, Biochem.* **55**: 631～661.
37. Roeder, G. S. et al. (1985) "The Origins of Gene Instability in Yeast." *Science* **209**: 1375～1380.
38. Boeke, J. D. et al. (1985) "Ty Elements Transpose Through an RNA Intermediate." *Cell* **40**: 491～500.
39. Mellor, J. et al. (1985) "A Retrovirus-Like Strategy for Expression of a Fusion Protein Encoded by Yeast Transposon Ty1". *Nature* **313**: 243～246.
40. Finnegan, D. J. (1985) "Transposable Elements in Eukaryotes." *Int. Rev. Cytol.* **93**: 281～326.
41. Mount, S. M., and Rubin, G. M. (1985) "Complete Nucleotide Sequence of the *Drosophila* Transposable Element *copia*: Homology Between *copia* and Retroviral Proteins." *Mol. Cell. Biol.* **5**: 1630～1638.
42. Rubin, G. M. (1983) "Dispersed Repetitive DNAs in *Drosophila*." in *Mobile Genetic Elements*, ed J. A. Shapiro, pp. 329～361, New York, Acad. Press.
43. Karess, R. E., and Rubin, G. M. (1984) "Analysis of P Transposable Elements Functions in *Drosophila*." *Cell* **38**: 135～146.
44. Fedoroff, N. (1983) "Controlling Elements in Maize." in *Mobile Genetic Elements*, ed. J. A. Shapiro, pp 10～63, New York, Acad. Press.
45. Pohlmann, R. F. et al. (1984) "The Nucleotide Sequence of the Maize Controlling Element *Activator*." *Cell* **37**: 635～643.
46. Temin, H. M., and Mizutani, S. (1970) "Viral RNA-Dependent DNA Polymerase." *Nature* **226**: 1211～1213.
47. Baltimore, D. (1970) "Viral RNA-Dependent DNA Polymerase." *Nature* **226**: 1209～1211.
48. Coffin, J. M. (1985) "Structure of the Retroviral Genome." in *RNA Tumor Viruses: Molecular Biology of Tumor Viruses*, 2nd ed., ed. R. Weiss et al., Cold Spring Harbor Lab, pp 261～368.
49. Bishop, J. M. (1983) "Cellular Oncogenes and Retroviruses." *Ann. Rev. Biochem.* **52**: 301～354.
50. Varmus, H. E. (1982) "Form and Function of Retroviral Proviruses." *Science* **216**: 812～821.
51. Panganiban, A. T. and Temin, H. M. (1984) "Circles with Two Tandem LTRs Are Precursors to Integrated Retrovirus DNAs." *Cell* **36**: 673～679.
52. Yamamoto, T. et al. (1980) "Indentification of a functional Promoter in the Long Terminal Repeat of Rous Sarcoma Virus." *Cell* **22**: 787～798.
53. Martin, G. S. (1970) "Rous Sarcoma Virus: A Function Required for the Maintenance of the Transformed State." Nature **227**: 1021～1023.
54. Takeya, T., and Hanafusa, H. (1983) "Structure and Sequence of the Cellular Gene Homologous to the RSV *src* Gene and the Mechanism for Generating the Transforming Virus." *Cell* **32**: 881～890.
55. Hunter, T. (1984) "The Proteins of Oncogenes." *Sci. Amer.* **251**: 70～79.
56. Bohmann, D. et al. (1987) "Human Proto-oncogene *c-jun* Encodes a Protein with Antigenic and Enhancer Bin-

ding Properties of Transcription Factor AP-1." *Science* **238**: 1386～1392.
57. Blair, D. G. et. al. (1981) "Activation of the Transforming Potential of a normal Cell Sequence: A Model for Oncogenesis." *Science* **212**: 941～943.
58. Chang E. M. et al. (1982) "Tumorigenic Transformation of Mammalian Cells Induced by a normal Human Gene omologous to the Oncogence of Harvey Murine Sarcoma Virus." *Nature* **297**: 479～483.
59. Shalloway, D. et al. (1984) "Overexpression of the c-src Protein Does Not Induce Transformation of NIH 3T3 Cells." *Proc. Nat. Acad. Sci.* **81**: 7071～7075.
60. Miller, A. D. et al. (1984). " c-*fos* Protein can Induce Cellular Transformation: A Novel Mechanism of Activation of a Cellular Oncogene." *Cell* **36**: 51～60.
61. Green, M. R. (1989) "When the Products of Oncogenes and Anti-Oncogenes Meet." *Cell* **56**: 1～3.
62. Fink, G. R. (1987) "Pseudogenes in Yeast? " *Cell* **49**: 5～6.

# 第十二章 基因工程导论[1~14]

## 一、概 述

自从1953年Watson和Crick阐明了DNA的双螺旋结构以来，生物学的发展进入了一个全新的阶段，即从整体和细胞水平跨入了分子水平。1973年美国首先成功地将基因在体外重组并通过质粒转入细菌内，进行无性繁殖，基因工程从此应运而生。基因工程或称基因重组技术，是将不同生物的基因在体外人工剪切组合并和载体(质粒，噬菌体、病毒)DNA连接。然后转入微生物或细胞内，进行扩增，并使转入的基因在细胞内表达，产生所需要的蛋白质。基因工程自70年代初产生后，发展极快，到70年代末已成功地用微生物来生产人的胰岛素和干扰素等医药品，80年代已将这一技术应用在高等生物的物种改造和新品种的培育上。由于这一技术的成功，人类发展到今天，已经进入改造和创建新的生命形态的时代。现将促进基因工程发展的几个重大发现按时间顺序列于表1。

**表1 促进基因工程发展的几个重大发现**

| | |
|---|---|
| 1869 | Miescher 首次分离 DNA |
| 1944 | Avery 证明细菌转化时，携带遗传信息的是DNA而不是蛋白质 |
| 1953 | Watson 和 Crick 在X光衍射实验的基础上提出 DNA 的双螺旋结构模型 |
| 1961 | Marmur 和 Doty 发现 DNA 的复性，并由此建立了核酸分子间杂交的可行性和专一性。 |
| 1962 | Arber 首次证明 DNA 限制性内切酶的存在，这导致 Nathans 和 H.Smith 纯化限制性内切酶，并用在 DNA 的序列分析上。 |
| 1966 | Nirenberg, Ochoa, 和 Khorana 破释了遗传密码。Gellert 发现了 DNA 连接酶。 |
| 1972～1973 | Boyer, Cohen 和 Berg 发展成功 DNA 克隆技术。 |
| 1975～1977 | Sanger 和 Gilbert 发展了DNA 序列快速测定法。 |

## 二、限制性内切酶对DNA分子的专一性切割

多种细菌能合成限制性核酸酶，这是保护它们自己，降解外来DNA分子的重要手段。每一种内切酶能识别DNA分子中由4～6个核苷酸组成的特定序列。而细菌细胞内的DNA则由于相应序列上的A或C碱基的甲基化而不被攻击，可是外源DNA一旦进入细胞立即被识别，双股DNA螺旋都被切断(图12-1)。目前已有100种以上具有不同专一性的限制性内切酶被提纯并已商品化。

特定的一种限制性内切酶能将DNA双股螺旋切割成一系列片段。用几种不同的限制性内切酶切割某一特定的基因，可构建出基因的限制性内切酶图谱(图12-2)，由于这种图谱反映出特定的核苷酸序列在基因中的排列位置，因此比较相关基因的限制性内切酶图谱可以大致看出相关基因间的同源性。图12-2是人和其他五种类人猿的血红蛋白基因的限制性内切酶图谱。自500万年前这些生物分化以来，它们的血红蛋白基因大致保持不变。

许多限制性内切酶都产生交错的切口，亦即在每条链的一端留下短的单链末端，即所谓的粘性末端，因为它们能和用相同内切酶切割生成的末端形成互补的碱基对。限制性内切酶能

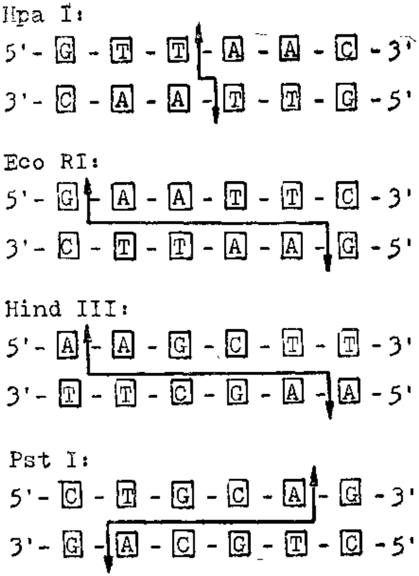



图 12-1 限制性内切酶能在特定的序列上切断双链 DNA

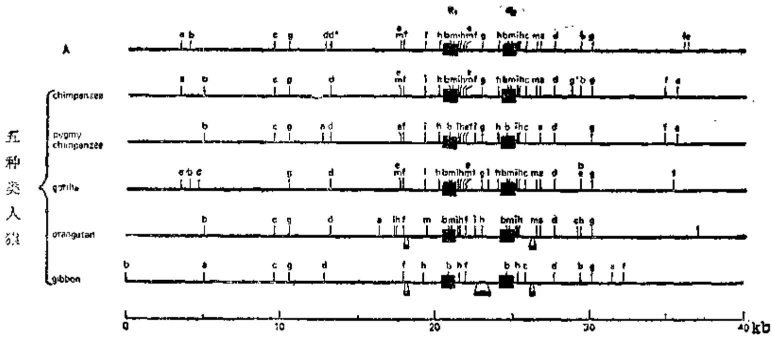



图 12-2 人类和其他五种类人猿的血红蛋白基因的限制性内切酶图谱

够产生粘性末端，在 DNA 重组技术中十分重要，因为它们能使任何两个 DNA 片断连结起来，只要这两个片段是用相同的内切酶或能产生相同粘性末端的两个内切酶切割的（详见第四章）。一旦两个粘性末端通过碱基互补相连后，就可用 DNA 连接酶封口，DNA 连接酶可在 DNA 的 5′-磷酸末端和 3′羟基末端之间形成共价磷酸二酯键（图 12-3）。

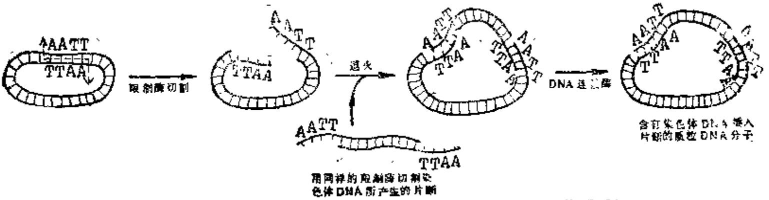



图 12-3　利用限制性内切酶和 DNA 连接酶可将染色体 DNA 片段重组到质粒 DNA 中去。

## 三、载体：克隆载体、穿梭载体和表达载体

进行任何一个 DNA 片段的克隆都需要有适合的载体。这种载体或者是质粒，或者是噬菌体，或者是病毒，通常大多数是经人工改造过的。作为载体必须具备两个条件：

(1) 该载体在细胞内必须能够进行自主复制，也就是说必须具备复制原点；

(2) 该载体必须具备适合的酶切位点，而且这些酶切位点不在复制原点区域内。这两个条件保证了载体的可繁殖性和可利用性。对于阳性克隆，尽管可以用放射性探针杂交的方法加以检测，但是人们常常还希望载体具有其他的筛选指标，如对抗菌素的抗性，某种基因产物

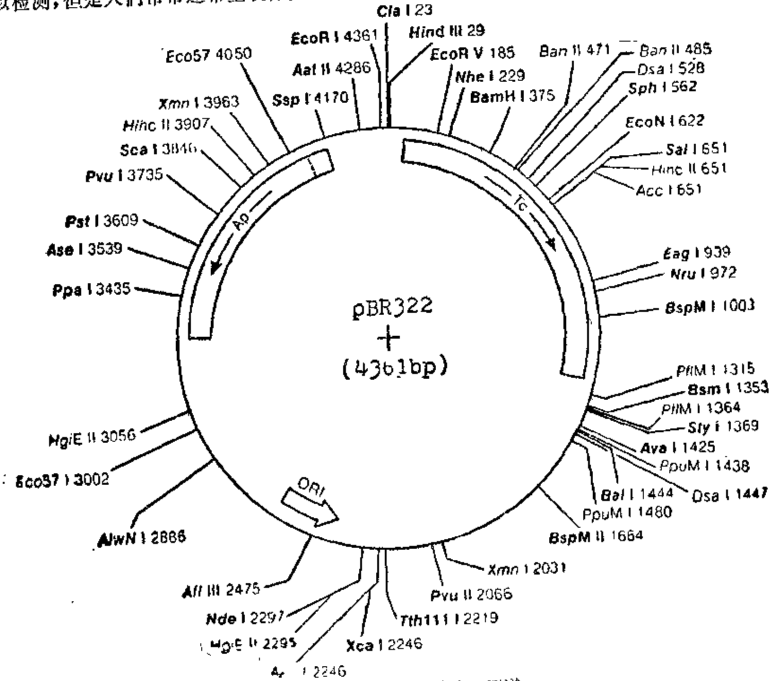



图 12-4a　pBR322 的酶切图谱

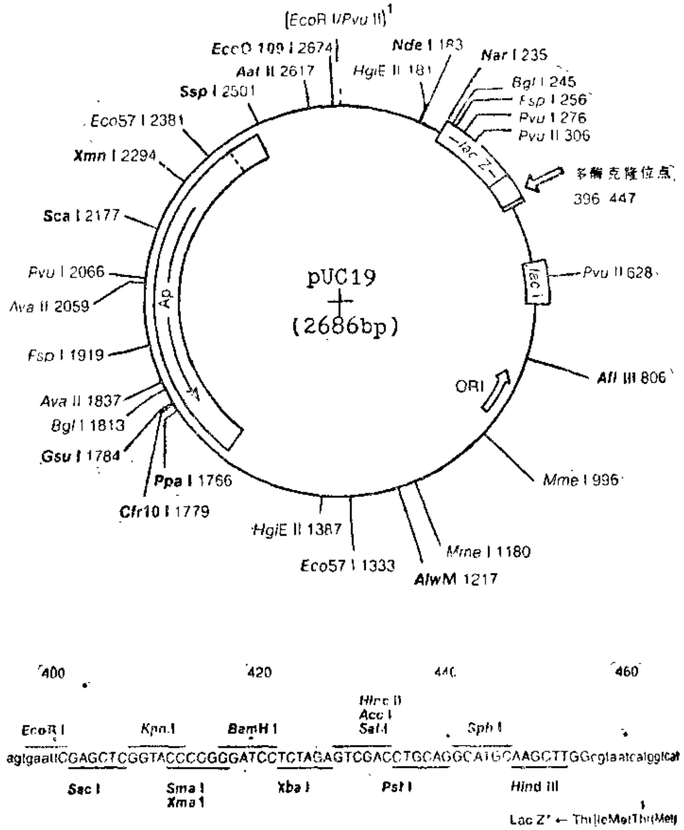



图 12-4 b　pUC 19 的酶切图谱

的显色反应等多重指标，以便使用过程中得心应手。

以繁殖 DNA 片断为目的的载体通常称为克隆载体。这种载体一般较小，在细胞内的拷贝数较高。也就是说，这种载体的复制不是处于寄主的严谨控制之下，而是处于松弛控制之下（每细胞内 10～200 个拷贝）。如果用最终浓度 10～20 $\mu$g/ml 的氯霉素停止细胞的蛋白质合成，则寄主 DNA 复制停止，而质粒的复制继续进行，使得每细胞内的拷贝数迅速增加，有时可达上千拷贝。这种处理方法称为质粒的扩增。由 Bolivar 等人于 1977 年构建的使用最为广泛的质粒 pBR 322 就是这样的一种载体(图 12-4a)。它有 4 361 bp，具有质粒 colE1 的复制原点，具有氨苄青霉素抗性基因和四环素抗性基因，因此筛选阳性克隆比较方便。现在世界上使用的绝大部分质粒载体均为 pBR 322 的衍生物。

1977年以来，Messing等人根据$\alpha$-互补原理构建了一系列利用显色法直接筛选的载体，这

就是pUC系列，如 pUC 8，pUC9，pUC 18，pUC 19 等 (图 12-4b)，还有相应的M 13 系列，如 M 13 mp8，M 13 mp9，M 13 mp18，M 13 mp19 等。什么叫 α- 互补呢？人们发现 *lacZ* 基因的突变体 M 15 (缺失氨基酸残基 11～41) 是没有野生型 *lacZ* 基因产物那种分解 X-gal 能力的。但是如果在 M 15 突变体的抽提物中加入 β- 半乳糖苷酶的第 3 至第 92 氨基酸残基的肽段 (α 肽) 则能恢复 M 15 分解 X-gal 而显示蓝色的能力。这样一种基因内互补现象称为 α- 互补。现在已经清楚，α 肽只需保持第 9 至第 59 个氨基酸残基就具有这种基因内互补能力。α-互补的发现，给设计 pUC 系列的载体带来极大的方便，载体中就没有必要包含 3 kb 以上的β-半乳糖苷酶的编码序列，而只要包含 lac 启动子和操纵子以及α肽编码序列再加上宿主菌是 *lacZ* M15 就行了。除此之外，为了保证质粒的稳定性，宿主菌还必须是 $lacI^q$，即*lac* 阻遏蛋白过量表达型。这样，只有在用 IPTG 诱导时，*lac* 系统才能够表达。当 pUC 系列载体内插入 DNA 片断时，由于破坏了 α 肽的阅读框或者引入了转录或翻译的终止信号，阻碍了 α 肽的表达，因而在加入IPTG 和 X-gal 后不能显示蓝色。pUC 系列载体仍然保留了一个氨苄青霉素抗性基因。M 13 系列的载体由于能够产生单链环 DNA，是进行序列分析和定点突变的重要载体。

特定的 DNA 序列可以通过克隆技术大量制造。任何来源的一段 DNA 都可以插入到载体中，然后转化到微生物细胞中扩增几百万倍。这一过程称为 DNA 克隆。其大致步骤是：纯化克隆载体 DNA，用限制性内切酶切割，然后与待克隆 DNA 片断退火以生成杂化质粒。最后将杂化质粒导入细菌中。最常用的导入方法是转化。在这一过程中，关键的步骤是用氯化钙溶液对细胞进行处理以增加细胞的通透性。经氯化钙处理过的细胞称为感受态细胞。只有少部分感受态细胞可以接纳质粒。阳性克隆可以通过抗菌素选择性培养或其它指标(如显色反应，放射性探针等)进行筛选。当转化的细菌分裂时，转入的质粒也复制，从而大量产生包含外源DNA的杂化质粒(图 12-5)。将杂化质粒DNA提纯，用克隆时所使用的限制性内切酶进行切割，通常就可以获得大量的外源性DNA 片断。

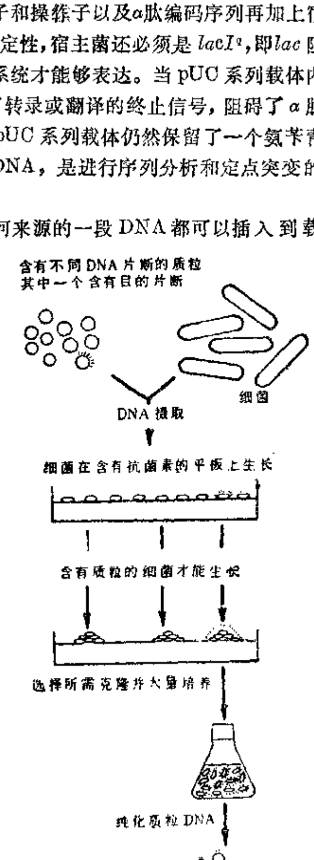


图 12-5 利用 DNA 克隆技术可以在细菌或酵母中大量扩增特定的DNA序列

穿梭载体(shuttle vectors)通常是指那些既能在真核细胞(哺乳动物细胞，酵母)中繁殖又能在原核细胞(如大肠杆菌)中繁殖的载体。因此这类载体必须既含有细菌的复制原点或细菌质粒的复制原点，又含有真核生物的复制原点，如 SV40 的复制原点或酵母的自主复制序列 ARS (autonomously replicating sequence)。除上述的两类复制原点之外，穿梭载体同样必须具备可资利用的酶切位点和适合的筛选指标。这样，穿梭质粒既可以用来转化细菌，又可以用来转化真核细胞。通常是将载体和待克隆的真核生物 DNA 片断先在细菌中进行繁殖、连接、再繁殖，最后转化到真核细胞中。

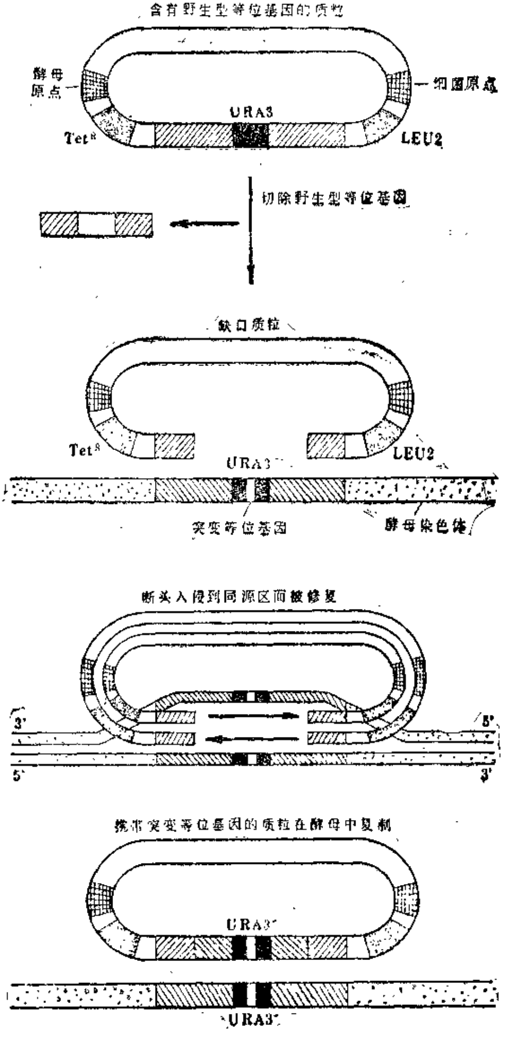



图 12-6 回收质粒的工作原理。本图所示是为了要回收酵母染色体中的突变基因 *ura*3⁻ 以进行序列分析，回收质粒中包含 *ura*3⁻ 基因及其两翼的同源序列。详细过程如正文所述。

穿梭载体不但可以将目的基因或 DNA 片断引入真核细胞，而且可以从真核细胞染色体上回收基因或 DNA 片断。在酵母细胞内，双链断裂的载体(断裂端)与染色体的同源部位之间具有很高的重组活性。利用这一原理，可将穿梭载体改造为回收载体 (retriever vectors)。在回收载体中除了穿梭载体所必需的成分之外，还必须带有待回收基因的同源拷贝或待回收基因两侧的同源拷贝。当回收载体构建完成后，即用限制性内切酶将载体上的待回收基因的同源拷贝切除。将这种缺口载体加到含有待回收基因的细胞中。当它一旦进入细胞核，具有很高重组活性的末端就侵入待回收基因的两翼序列中。在载体末端的 3′-OH 上就以待回收基因为模板，制造出待回收基因的新拷贝。这种载体和染色体的重组产物经过拆分，产生了具有繁殖能力的载体并以其所携带某种基因(如亮氨酸合成酶基因)与寄主的缺陷型基因互补(未经修复的缺口载体不具备此种能力)。分离纯化这种已有待回收基因的载体，再穿梭到 *E.coli* 中去扩增并进行序列分析。图 12-6 就是这种回收载体的工作原理。

表达载体是用来将克隆的外源基因在寄主细胞内表达成蛋白质的载体。作为表达载体的必要条件是：(1) 很强的启动子，能为寄主的 RNA 聚合酶所识别；(2) 很强的终止子，以便使 RNA 聚合酶集中力量转录克隆的基因而不去转录其它无关的序列，同时很强的终止子所产生的 mRNA 较为稳定(见第八章)；(3) 启动子是必须受控制的，只有当诱导时才能进行转录。这一点也十分重要，因为不受控的转录不但效率不高(见第十章)，而且给细胞造成不利影响，特别是当克隆的基因产物对细胞有害时更是如此。(4) 所产生的 mRNA 必须具有翻译的起始信号，即 AUG 和 S-D 序列。当然作为任何种类载体所必需的复制原点和灵活的酶切位点也是表达载体不可缺少的重要成分。表达载体的启动子种类很多，这里不能详细介绍。使用得最多的启动子乃是λ噬菌体的 $p_L$ 和 $p_R$ 启动子。$p_L$ 和 $p_R$ 不但启动子强度高，而且受 *CI* 基因产物的严格控制。生理过程中 $p_L$ 和 $p_R$ 的除阻遏作用是十分复杂的(见第八章)。在基因工程中，利用 $CI^{857}$ 温度敏感突变体却十分方便。突变基因 $CI^{857}$ 或者在载体上，或者在寄主的染色体上(二者各有长处)。当细胞在 30℃ 生长时，$CI^{857}$ 阻遏蛋白牢牢地阻遏 $p_L$ 启动子，把温度提高至 40～42℃ 时，则立即除阻遏，从 $p_L$ 或 $p_R$ 进行大量的转录并翻译。PPL-lambda 及其衍生载体和 pEX 系列载体可以作为这类表达载体的代表(图 12-7a)。

乳糖操纵元的启动子，尽管强度不算很强，但由于有灵敏的诱导方法和检测方法(见上述)，也在广泛地使用。特别是人们将之改造为杂合启动子，如利用 *trp* 启动子的 －35 序列和 *lacUV*－5 的 －10 序列相结合而成为很强的杂合启动子 *tac*。*pKK* 系列载体就是这样的启动子(图 12-7b)。

表达载体根据产生的蛋白序列，又可分为融合型载体和非融合型载体。尽管人们大多希望得到非融合蛋白质，但是融合型载体也有其优越性。因为融合蛋白质的 N 端部分通常为细胞内本来就有的成分，使得克隆基因的产物较为稳定。另一方面，AUG 后大约 20 个核苷酸对翻译效率也有影响，而载体中的编码序列(融合蛋白 N 端部分)大多是表达体系中天然存在的，翻译效率较高。还有，利用表达体系中的信号肽序列有时能得到分泌型的表达。有时，用较小的肽段作为抗原时，必须使之成为融合蛋白。再说，表达出来的融合蛋白可以通过化学方法裂解产生所需的非融合蛋白。尽管如此，不到不得已，人们还是宁愿采用非融合型表达载体。

## 四、基因组克隆和 cDNA 克隆

用限制性内切酶切割细胞的整个基因组 DNA，可以得到大量的基因组 DNA 片断。然后

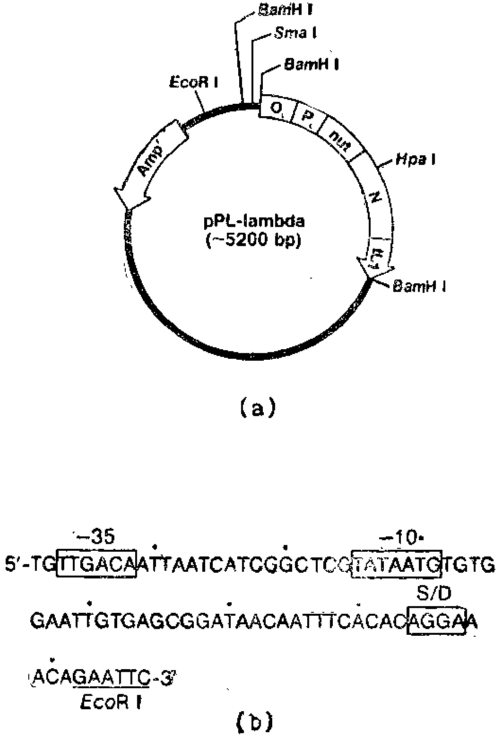



图 12-7 (a) 含有 $p_L$ 启动子的质粒 pPL-lambda；
(b) 质粒 pKK223-3中的杂合启动子 *Ptac* 及其附近序列

将这些 DNA 片断与载体连接，再转化到细菌中去，让宿主菌长成克隆。这样，一个克隆内的每个细胞的载体上都包含有特定的基因组 DNA 片断。这样的一套克隆就叫做基因组克隆；其中克隆的一套基因组 DNA 片断就叫做基因组文库。为了避免某一个基因被分割于两个以上的克隆中，常常在制作基因组文库时使用不完全酶切法。随着要求不同，可使用不同的载体制作基因组文库。pBR 322 一般可容纳 5kb 的外源 DNA，λ噬菌体可容纳 22kb，而 cos 质粒(cosmid)则可容纳 40 至 50 kb。这样，酵母基因组约 14 000 kb，可以容纳于 3 000 个 pBR 322，或 1000 个λ噬菌体，或 300 个 cos 质粒。如果用λ噬菌体作为载体，人类基因组文库需要 250 000 个不同的克隆。如果要求的精确度高，所需的克隆数就高达数百万个。如何从众多的基因组克隆中筛选出少数含有目的基因的克隆，是 DNA 重组技术中最困难的工作。通常用来鉴别特定的 DNA 片断的方法是用放射性标记的探针，这种探针能够与克隆化的 DNA 片断杂交。

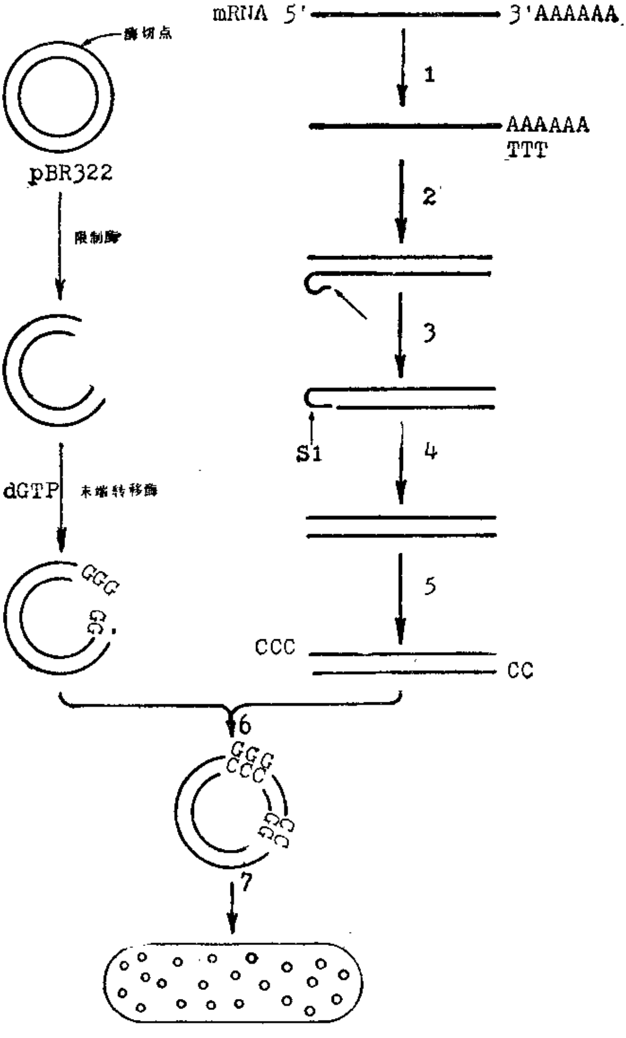



图 12-8 cDNA 克隆的制作过程：

(1) 寡聚T加入到 poly(A) mRNA 中；(2) 逆转录酶进行逆转录；(3) 用 NaOH 降解mRNA，DNA 聚合酶 I 以发夹末端为引物合成 cDNA 的互补链；(4) 用单链特异性核酸酶 S1 降解发夹噜噗；(5) 用末端 DNA 转移酶加上寡聚 C；(6) 与末端带有寡聚 G 的开环质粒退火；(7) 转化大肠杆菌并扩增。

上面所述的克隆方法常被称做鸟枪法(shotgun)，因为整个基因组DNA被切割成许多片断，其结果是某些片段含有基因或部份基因，某些片断则只含有非编码DNA。另一种方法是选择性地克隆那些转录成mRNA的DNA序列。这种方法首先从细胞中提取mRNA，然后通过逆转录酶合成mRNA的DNA拷贝，即cDNA。单链cDNA分子再通过DNA聚合酶的作用转变成双链cDNA分子，然后再插入质粒中并克隆化(图12-8)。

已有各种方法构建成特别的质粒，这种质粒可使克隆化的cDNA在细胞内合成大量特定的蛋白质，利用这种基因工程手段已使细菌或酵母大量产生人胰岛素，生长激素和干扰素。

cDNA也可用来鉴定用鸟枪法得到的含目的基因的克隆。通常用放射性核苷酸单体合成单链cDNA。这种放射标记的DNA可和基因组克隆互补杂交，因为cDNA是以mRNA为模板合成，它必定相应于编码某一蛋白质的基因组DNA序列，和这种探针杂交的DNA克隆必然是含有编码mRNA分子的基因。

### 五、核苷酸链的杂交反应提供了检测特定核苷酸片段的灵敏方法

将DNA水溶液在100℃下加热，或高pH(pH≥13)处理，维系DNA双股螺旋的互补碱基对将被破坏，双股螺旋分离成单链，这个过程称DNA的变性。人们在相当长的一段时间内都认为这是一个不可逆发应。1961年证明互补的单股DNA链在65℃下长时间保温可重新形成双股螺旋。这种链间的杂交可发生在任何两条单股核酸链之间：DNA-DNA，DNA-RNA或RNA-RNA，只要它们具有互补的核苷酸序列。

由于双螺旋形成的速度受限于两条互补核苷酸链之间相互碰撞的几率，因此携带特定核苷酸序列的DNA浓度可根据它和放射标记的互补DNA探针间的杂交速度来测定。这是一种很灵敏的测定方法，即使互补序列的浓度低到每个细胞只有一个分子也能检测出来，这种方法已被用来检测细胞中某一特定DNA序列的拷贝数。通常的基因在单倍体基因组中只有一个或几个拷贝，而重复序列则可能有成千上万个拷贝。

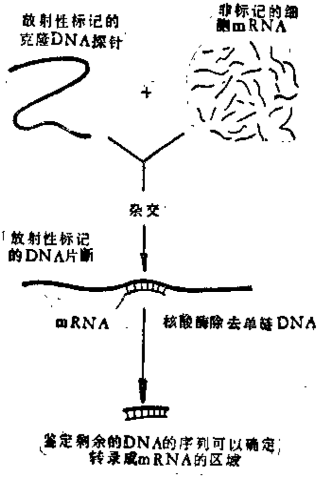



图12-9 用DNA·RNA杂交技术探明基因转录成mRNA的区域。用此技术发现了真核生物的内元。

分子杂交也被用来检测克隆的DNA是否转录成RNA，每个细胞中含有几个RNA拷贝，以及探针和细胞中RNA杂交的精确部位，从而探明RNA转录的起始和终止部位(图12-9)。

放射标记的DNA探针已被广泛用来检测经电泳分离的DNA混合片段中某一特定的DNA序列。经电泳分离的DNA凝胶可用墨迹法转移到硝酸纤维纸上，与放射标记的DNA探针杂交的片段可用放射自显影检定出来(图12-10)。用类似的方法，也可将生长在琼脂培养基上的菌落做成复制皿后再转移到硝酸纤维纸上变性并与特定的放射标记的探针杂交，藉此也可以筛选出带有特定DNA片段的菌落。这种技术称为菌落杂交，又称为原位杂交。

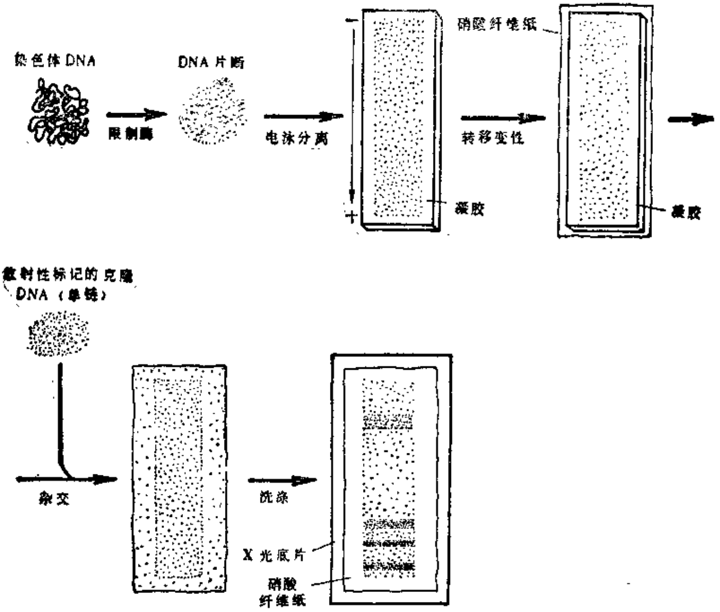



图 12-10 用放射性标记的 DNA 探针寻找特定的染色体DNA片断。多余的探针被洗去，能与探针杂交的 DNA 将在X光底片上显影成带，

## 六、用 DNA 重组技术大量制造细胞中的微量蛋白质

在 DNA 重组技术出现以前，只有在细胞中含量较多的蛋白质才可能纯化并加以研究。含量占细胞蛋白总量 10% 的蛋白质可以用几百克细胞为原料，通过层析，电泳等一系列纯化步骤，有可能得到 100 毫克纯蛋白质，这个量足够完成氨基酸序列分析，生物活性分析以及制备抗体，藉此可研究这种蛋白质在细胞中的定位。如果能长出合适的晶体，则可用 X 衍射分析这个蛋白质的三维空间结构。许多含量较大的蛋白质如血红蛋白，胰蛋白酶，免疫球蛋白和溶菌酶等，就是这样解决了它们的结构和功能。但是在真核细胞中的绝大多数蛋白质，其中包括重要生物功能的蛋白质在细胞中含量极微，用通常的方法即使得到几微克纯样品也是十分困难的，甚至是不可能的。但是用重组 DNA 技术，原则上任何一种蛋白质都有可能大量制备。采用的典型步骤总结如下：

1．用层析和高分辨电泳技术纯化微克量的目的蛋白质。

2．用微量顺序仪测定该蛋白质的 N- 端 30 个氨基酸碱基序列。

3．根据遗传密码预测相应于氨基酸序列的 mRNA 核苷酸序列。用快速固相化学合成法合成一系列短的 DNA 片段，15～20 核苷酸长，它们能和 mRNA 的部份序列形成互补碱基对。

4. 将这些短的 DNA 片段和细胞总的 mRNA 杂交，藉此可分离出所需要的 mRNA，并通过反转录酶合成 cDNA。

5. 通过克隆技术扩增所需的 cDNA。

6. 扩增的 cDNA 和细胞总的 mRNA 杂交，分离并纯化较多量的 mRNA，并将此 mRNA 进行体外翻译以验证是否为目的基因。

7. 将克隆化的 cDNA 序列插入特定的质粒 DNA 中，这种载体含有转录和翻译的起始与终止信号，将载体质粒转入细菌或酵母细胞中进行表达。

## 七、植物的基因工程

传统的植物育种和原生质体融合都是基因操作方法，可用来改变植物的遗传性状，但是这些方法在基因转移上都缺乏像重组 DNA 技术用于微生物那样的精确性。植物遗传工程的目的往往是要把单个特定的基因导入已经被广泛利用的各种植物中去，这就涉及两个基本问题：首先必需得到一定量的特定的纯基因，其次，设计合适的方法务使这些基因能插入植物的染色体中并表达。第一个问题按目前已广泛应用的重组 DNA 技术，可将任意一段 DNA 在细菌细胞中扩增，应该没有不能克服的困难。麻烦在于如何从一大堆克隆中筛选出所需要的基因，特别是那些影响诸如产量的基因，高产性状可能是受多基因控制的，其生物化学机理尚不清楚。第二个问题，即如何将外源基因导入植物中也还存在许多尚未解决的问题。近年来关于 Ti 质粒的研究使人们开始摸索出解决植物基因工程中的载体难题。

土壤微生物土壤杆菌 (*Agrobacteria*) 中的某些属能诱发某些双子叶植物产生根瘤 )crown gall)，植物的根瘤细胞和动物的癌细胞类似，它们能独立地、不受控制地生长。土壤杆菌导致发生根瘤的原因是由于在这种细菌中存在一种质粒，当这种质粒的 DNA 整合到宿主植物细胞的染色体中就诱发出根瘤，并使瘤细胞合成 opines。这种质粒被称为 Ti 质粒。Ti 质粒是环状 DNA 分子，分子量约为 $1.2\times10^8$，它们以独立的复制单位存在于细菌中。研究证明，Ti质粒至少具有三种生物学功能，即诱发根瘤，opine 化合物的合成和抑制细胞分化。负责这三种功能的基因在 Ti 质粒 DNA 中靠得很近，因此包含这三个基因的 DNA 片段被称为转化 DNA 或 T DNA。用 Southern 墨迹法已证明，Ti 质粒的约 20 000 个碱基对完全共价整合到根瘤细胞的 DNA 中。

Ti 质粒由于具有以下性质，因此很适合作为导入外源基因的载体。首先，土壤杆菌的寄生范围很广，实际上它们几乎可以使所有双子叶植物的细胞转化。其次，整合到植物细胞中的 T DNA 按孟德尔规律遗传，而且基因具有自己启动子，因此可以偶联外源基因并使之表达。导入 T DNA 的最简单的方法是用含有 Ti 质粒的土壤杆菌感染植物细胞，其余的一切让它自然完成，外源基因必需插入到Ti 质粒的T DNA序列中。因为整个 Ti 质粒比通常用的质粒大得多，因此可先用限制性内切酶将 T DNA 切割出来，并插入到常用的PBR 322质粒中，使之在大肠杆菌中扩增并分离，然后再用 DNA 重组技术将目的基因插入到 TDNA 中。这种杂化的 TDNA 在大肠杆菌中大量扩增后再转入土壤杆菌中，并和其中的天然 Ti 质粒发生同源重组，结果将带有目的基因的 T DNA 转移到 Ti 质粒中并取代正常的 T DNA。这种带有外源基因的 Ti 质粒通过土壤杆菌再去感染植物，最终使所需的外源基因导入植物中（图 12-11）。

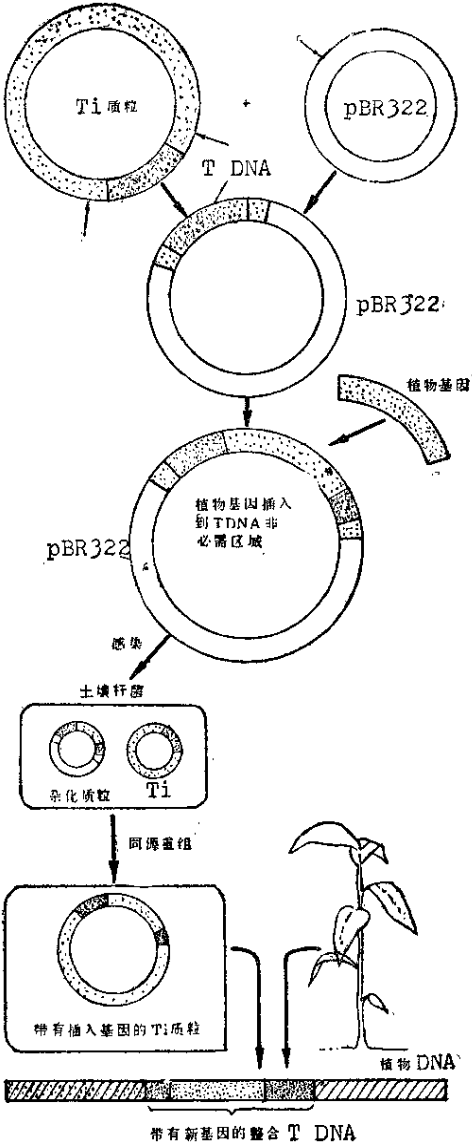



图 12-11 用 Ti 质粒作为载体进行植物基因工程的步骤。首先用限制酶将 Ti 质粒的 T DNA 片断切出来并克隆到质粒 pBR 322 中。其次将目的基因插入到 T DNA的非必需区域。然后将杂种质粒与含有正常 Ti 质粒的根瘤杆菌混合。两种质粒的 T DNA 发生同源重组。外源基因就能进入 Ti 质粒。再用含有这种 Ti质粒的根瘤杆菌感染植物，就可将外源基因整合到植物染色体上。

## 八、基因工程工业

分子生物学家成功地开发克隆技术的同时，他们已经认识到这种技术在商业上的潜在价值和在科学上的潜在价值同样重要。基因工程工业的目标是利用克隆技术更有效、更便宜地制造出有用的产物，主要是蛋白质。例如胰岛素、干扰素，纤溶酶原激活剂和第Ⅷ凝血因子等基因工程产品都已问世。利用 DNA 重组技术制造安全疫苗，特别是对那些不能用动物细胞培养的病毒，例如乙型肝炎病毒疫苗，更具价值。除制药工业外，工程菌在化学、石油甚至采矿工业中都有潜在应用价值。下面列举几个典型的例子。

**1. 胰岛素的基因工程** 胰岛素从核糖体合成出来时带有一个信号肽，输送过程中信号肽

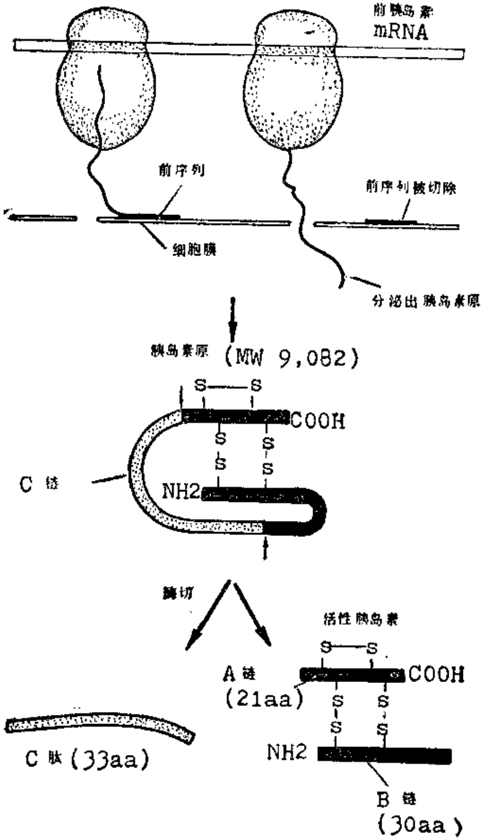



图 12-12 核糖体合成的前胰岛素原在分泌时切除"前"序列产生胰岛素原，再切除 C 肽产生胰岛素。

被切除，生成胰岛素原，胰岛素原是一条单链由三对二硫键连接弯曲成一环圈。具有生理功能的成熟胰岛素由两条链组成。A 链含 21 个氨基酸残基，B链含 30 个氨基酸残基 (图 12-12)。胰岛素原在细胞中被酶解，切去 35 个氨基酸残基的 C 肽而生成具有生理活性的胰岛素。因此胰岛素有三种形式：(1) 前胰岛素原是翻译的直接产物，具有信号肽，C 肽和 A、B 链；(2) 胰岛素原已失去信号肽但仍保留 C 肽；(3) 胰岛素只有 A、B 链。

用基因工程生产胰岛素不宜从染色体取得基因开始，因为这样做最终产物不是胰岛素而是前胰岛素原，而且还含有不编码的介入序列。克服这一困难的方法是分别合成编码胰岛素 A、B 链的核苷酸序列，因为胰岛素的氨基酸序列已经清楚，因此 A 链需要 63 个核苷酸，B 链需要 90 个核苷酸，两条链的起始端都加甲硫氨酸的密码子 ATG，以便表达产物的后加工。合成的 A、B 链基因分别插入已装配在载体质粒上的 β-半乳糖苷酶的基因中，将此重组质粒转入大肠杆菌中，克隆化的基因将受 β-半乳糖苷酶基因启动子的控制，转录并表达胰岛素 A 链或 B 链和部分 β-半乳糖苷酶的融合蛋白。由于此融合蛋白中有一额外的甲硫氨酸残基，因此可用溴化氰裂解，将胰岛素A或 B 链与 β-半乳糖苷酶片段分开。利用甲硫氨酸作为融合蛋白的连结点只能用于目的蛋白不含甲硫氨酸或色氨酸。纯化的 A 链和B 链再用二硫桥连结起来得到完整的胰岛素 (图 12-13)。利用基因工程生产人胰岛素的另一种方法是将胰岛素原的 mRNA通过反转录酶合成cDNA。在合成的胰岛素原的cDNA的5′末端加上甲硫氨酸的密码子 (ATG)，将此 cDNA 插入载体质粒的 β-半乳糖苷酶基因中并转入大肠杆菌表达，和细菌半乳糖苷酶融合的胰岛素原可用 CNBr 处理而释放出来，胰岛素原自动褶叠形成天然状态的三维结构，二硫桥也自然形成，最后切去 C 肽而生成胰岛素(图 12-13)

从商业开发的角度考虑，理想的目标是工程菌单个细胞的产率高，而且能将所需的蛋白质分泌到培养介质中，这样就可以用连续培养大量获得目的蛋白。不需要收集和破碎细胞从而简化了后处理过程。解决这个问题的一个可行方法是将胰岛素基因插入质粒 β-内酰胺酶基因中，β-内酰胺酶是降解青霉素的酶，通常可由细菌分泌到介质中，将这种重组质粒转化到大肠杆菌中并使之扩增，细菌就将胰岛素原和 β-内酰胺酶融合蛋白分泌到介质中，β-内酰胺酶就作为一种载体蛋白(图 12-14)。

**2. 人生长激素的基因工程** 人生长激素有 191 个氨基酸残基组成的单链蛋白质激素，由脑下垂体分泌，侏儒症是由于缺乏生长激素所致。已成功地用酶促合成cDNA及部分化学合成生长激素 DNA 而得到生长激素的完整基因，其中编码氨基酸残基 1～24 的 DNA 片段是化学合成的。在第一个密码子前而加上 ATG 碱基以保证蛋白质合成的起始。将合成的基因插入表达质粒并导入大肠杆菌，从而使大肠杆菌制造人生长激素。细菌产生的生长激素多了一个起始密码子编码的甲硫氨基酸残基，其余部份和脑垂体分泌的人生长激素全同，生物活力也不受影响。

**3. 干扰素的基因工程** 和胰岛素、人生长激素不同，在干扰素重组 DNA 成功以前，对干扰素的结构一无所知，因此不可能人工合成基因。干扰素 cDNA的获得是将产生干扰素的白细胞的 mRNA 分级分离，然后将不同部份的 mRNA 注入蟾蜍的卵母细胞，并测定干扰素的抗病毒活性。结果发现12S mRNA的活性最高，因此用这部分mRNA合成cDNA。干扰素mRNA 仅占白细胞总 mRNA 的 0.01～0.1%，因此转化的 cDNA 文库中大肠杆菌克隆多达上万个，每一个克隆都需检测是否含有专一的干扰素基因。按照干扰素的 cDNA 序列推出干扰素的氨基酸序列。α 和 β-干扰素由 166 个氨基酸残基组成，但信号肽分别由 21 和 23 个氨基酸残基

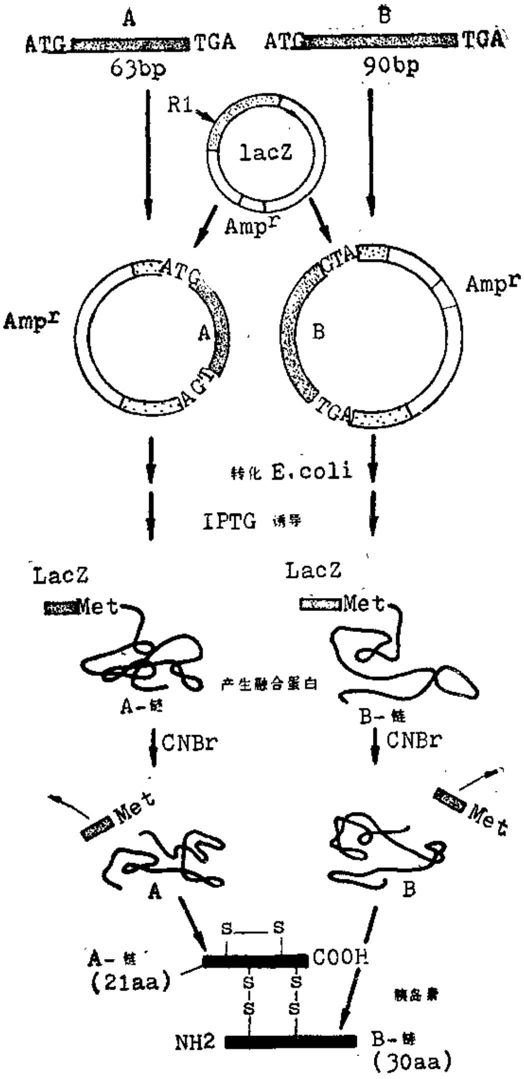



图 12-13　用合成的胰岛素基因在细菌中生产胰岛素，编码胰岛素 A、B 链的寡核苷酸系化学合成，ATG 加在 5′ 端。合成的基因克隆在 β- 半乳糖苷酶启动子的下游，仍保留部份 β- 半乳糖苷酶基因。转入 *E.coli* 后表达产生一杂合蛋白，由 β- 半乳糖苷酶的 N 端部份和胰岛素 A 链或 B 链组成，并由甲硫氨酸相连结。通过 CNBr 裂解可将 β- 半乳糖苷酶除去。

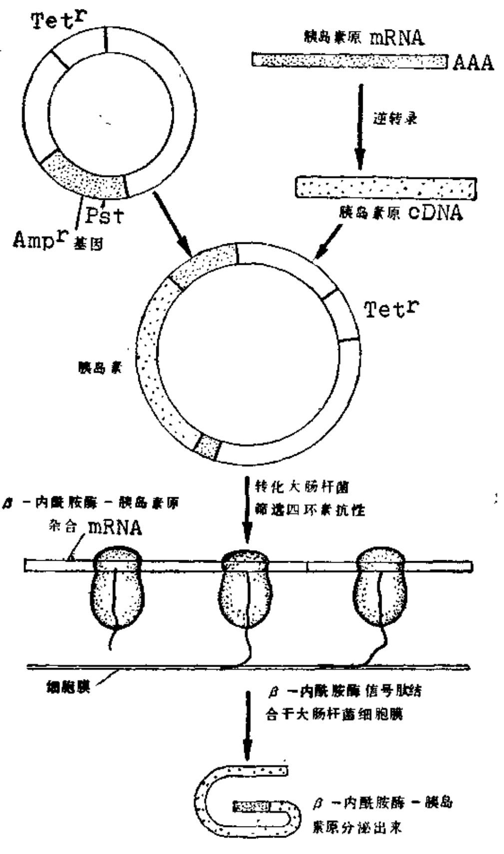



图 12-14 原核细胞的信号顺序使融合蛋白可以分泌出来。胰岛素原 cDNA 克隆在 β-内酰胺酶基因中，结果产生 β-内酰胺酶——胰岛素原融合蛋白。由于 β-内酰胺酶 N 端有信号肽，可使大肠杆菌分泌此融合蛋白。

组成。α-干扰素有13个亚型，分别由13个基因编码，而 β-干扰素只有一个基因。所有的人 α-干扰素基因和单个 β-干扰素基因都在第9染色体上。β-干扰素基因和 α-干扰素基因没有内元。大多数干扰素都是糖基化的，但糖链并非生物活性所必需。起初干扰素在大肠杆菌中的表达水平为每个细胞生产50个干扰素分子，将干扰素 cDNA 序列精确定位于 *E. coli* 的乳糖启动子邻近，提高表达效率1 000倍以上。目前1升培养液已可得到1mg干扰素，并已纯化结晶，证明即使大剂量应用也无毒性。γ-干扰素的基因工程困难更多，但也已克隆成功。用细菌毒素诱导人脾脏细胞并提取 mRNA 并注入蛙卵母细胞中，检定 mRNA 的翻译产物是否有 γ-干扰素活性，将具有 γ-干扰素的 18 S mRNA 逆转录为 cDNA，并将此 cDNA 克隆到具

有SV40复制原点和质粒复制原点的穿梭载体中，将重组质粒导入大肠杆菌中增殖再转入COS细胞。COS细胞是整合有SV 40大T抗原基因的猴细胞株。在COS细胞中SV40复制序列使载体大量扩增并使cDNA表达。带有干扰素cDNA的克隆可用来从人染色体基因库中筛选出干扰素基因。和α、β-干扰素不同，γ-干扰素基因有三个内元，在人基因组中是单拷贝。现在已设计出多种宿主载体系统来大量合成干扰素用以治疗病毒感染和癌症。

4. 乙型肝炎工程疫苗　基因克隆技术对那些不能用细胞培养生长的病毒疫苗最具重要意义。例如乙型肝炎是世界性广泛传播的严重传染病。由于乙型肝炎病毒不能用动物细胞培养生长，因此无法制造疫苗。在DNA重组技术出现以前，唯一的方法是从患者血液中分离病毒颗粒，灭活后作为疫苗，因此来源受到很大限制。1979年乙型肝炎病毒的基因组，一条部份单股的DNA，约3 200碱基对克隆成功并完成了序列测定，从而探明了2个已知的病毒蛋白的氨基酸序列，即病毒的核心蛋白(和病毒DNA结合在一起)和表面抗原蛋白。现已成功地将核心蛋白基因克隆到表达载体上并在大肠杆菌中大量表达，其产物已用于肝炎的诊断。表面抗原蛋白在大肠杆菌中的表达不很成功。最近将表面抗原基因和自主复制的酵母质粒的启动子连结并转入酵母细胞中，转化的酵母细胞能合成一种颗粒。它在抗原性和结构上都和22-nm病毒颗粒十分类似，这种颗粒实际上是中空病毒颗粒，只含有表面抗原蛋白和磷脂而没有核心蛋白，这种颗粒虽未糖基化，但并不影响其抗原性，因此可作为乙型肝炎疫苗。

## 九、结束语

人类研究自然科学的目的有两个，一个是了解自然界物质的运动规律，另一个是利用这些规律去改造自然。本书作为一门专业基础理论课程，当然其大部分篇幅是去描述前者；而对于后者，本书最后一章概括地展现人类改造生命与创造新生命的伟大业绩。现在基因工程的发展已进入新阶段，人们不再满足于制造生命体内所存在的生物活性大分子，而是要综合这些大分子的许多优点，制造出全新的性能更好的物质。这就是蛋白质工程，它是基因工程和蛋白质化学相结合的产物。人类是万物之灵，而我国人口占全人类的五分之一，应当对于这一伟大业绩作出更大的贡献。作者们正是怀着这种心情撰写该书，希望它能在分子遗传学的教学工作中发挥作用并为促进我国生命科学的发展作出些微贡献。

## 参　考　文　献

1. Watson, J. D., and Crick, F. H. C (1953) "Molecular Structure of Nucleic Acids: A Structure for Deoxyribose Nucleic Acid." *Nature* 171: 737～738.
2. Watson, J. D. et al. (1983) *Recombinant DNA: A Short Course*. Scientific American Books, New York.
3. Rodrigues, R. L. and Tait, R. C. (1983) *Recombinant DNA Techniques: An Introduction*. Addison-Wesley Publishing Company, London.
4. Maniatis, T. et al. (1982) *Molecular Cloning*. Cold Spring Harbor Laboratory, New York.
5. Silhavy, T. J. et al. (1984) *Experiments with Gene Fusions*. Cold Spring Harbor Laboratory.
6. Balba's, P. et al. (1986) "Plasmid Vector pBR322 and Its Special-Purpose Derivatives — A Review." *Gene* 50: 3～40.
7. Langley, K. E. et al. (1975) "Amino Acid Sequence of β-Galaclosidase. IV. Sequence of an α-Complementing Cyanogen Bromide Peptide, Residues 3 to 92." *J. Biol. Chem.* 250: 2587～2592.
8. Yanisch-Perron, C. et al. (1985) "Improved M13 Phage Cloning Vectors and Host Strains: Nucleotide Sequences of the M13mp18 and pUC19 Vectors." *Gene* 33: 103～119.
9. Remaut, E. et al. (1981) "Plasmid Vector for High-Efficiency Expression Controlled by the $P_L$ Promoter of

Coliphage Lambda." *Gene* 15: 81～93.

10. Derom, C. et al. (1982) "High-Level Synthesis in *Escherichia coli* of the SV40 Small-t Antigen Under Control of the Bacteriophage Lambda $P_L$ Promoter." *Gene* 17: 45～54.
11. Stanley, K. K. and Luzio, J. P.(1984) "Construction of a New Family of High Efficiency Bacterial Expression Vectors: Identification of cDNA Clones Coding for Human Liver Proteins." *EMBO J*. 3: 1429～1434.
12. Amann, E. et al. (1983) "Vectors Bearing a Hybrid trp-lac Promoter Useful for Regulated Expression of Cloned Genes in *Escherichia coli*." *Gene* 25, 167～178.
13. Bolivar, F. et al. (1977) "Construction and Characterization of New Cloning Vehicles. II. A Multipurpose Cloning System." *Gene* 2: 95～113.
14. Hohn, B., and Collins, J. (1980) "A Small Cosmid for Efficient Cloning of Large DNA Fragments." *Gene* 11: 241～298.
15. Gluzman, Y. (1982) *Eukaryotic Viral Vectors*. Cold Spring Harbor Laboratory, New York.
16. Krawetz, S. A. et al. (1986) "Laboratory Methods: A Vector-Primer-Cloner-Sequencer Plasmid for the Construction of cDNA Libraries." *DNA* 5: 427～435.
17. Messing, J. et al. (1981) "A System for Shotgun DNA Sequencing." *Nucleic Acid Res*. 9: 309～321.
18. Shortle, D. et al. (1981) "Directed Mutagenesis." *Ann. Rev. Genetics* 15: 265～294.
19. Chambon, P. (1981) "Split Genes." *Sci. Amer*. 244: 60～71.
20. Hooykaas, P. J. J. and Schilperoort, R. A. (1985) "The Ti-plasmid of *Agrobacterium tumefaciens*: A Natural Genetic Engineer." *TIBS* 10: 307～309.
21. Yanofsky, M. F. et al. (1986) "The *virD* Operon of *Agrobacterium tumefaciens* Encodes a Site-specific Endonuclease." *Cell* 47:471～477.
22. Stachel, S. E. and Zambryski, P. C. (1986) "*virA* and *virG* Control the Plant-induced Activation of the T-DNA Transfer Process of *A. tumefaciens*" *Cell* 46: 325～333.
23. Szoka, P. R. et al. (1986) "A General Method for Retrieving the Components of a Genetically Engineered Fusion Protein." DNA 5: 11～20.
24. Bone, R. et al. (1989) "Structural Plasticity Broadens the Specificity of an Engineered Protease." *Nature* 339: 191～195.

# 附录：涵义和译名有所变化的部份分子遗传学术语

1．Gene——基因 1909 年，丹麦遗传学家 W. Johannsen 首先使用基因一词，其概念与当年孟德尔提出的“遗传因子”完全一致。这一阶段，基因只是逻辑推理的产物，用作表示生物性状的符号，并无实质内容。

后来，T. H. Morgan 等人依据对果蝇的研究，提出连锁交换规律和伴性遗传，指出基因在染色体上直线排列。1926 年 Morgan 发表了《基因论》并指出基因代表着一个有机的化学实体。

自 Morgan 以后，遗传学家一直在探寻基因的化学实体。1944 年，O. T. Avery 等人做了肺炎球菌的转化实验，才首次证实基因的本质是 DNA，是 DNA 分子上的功能单位。1953 年，J. D. Watson 和 F. H. C. Crick 提出了 DNA 结构的右手双螺旋模型。从此，生物化学家和遗传学家开始向基因大进军，以探明基因的结构、表达和调控的细节，便诞生了分子遗传学。

1955 年，S. Benzer 研究了 T4 噬菌体 rⅡ 区的精细结构，提出了顺反子(cistron)，突变子(muton)和重组子(recon)等术语。顺反子学说认为顺反子是一段编码一种多肽链的核苷酸序列，是一个功能单位，如 rⅡ 区的 A 亚区和 B 亚区是两个顺反子，编码两种多肽链。而突变子和重组子经后来证明实际上是一个核苷酸对。顺反子学说彻底否定了基因是决定遗传性状的功能单位和突变、重组的最小单位这样三位一体的概念。顺反子学说认为，一个顺反子就是一个基因，这个基因或者编码蛋白质，或者编码 RNA 分子(tRNAs, rRNAs)。

1961 年，F. Jacob 和 J.Monod 提出了操纵元(又称操纵子)的概念，揭示了原核生物基因表达调控的重要规律。在他们提出的乳糖操纵元模型中，将基因分为结构基因，调节基因和所谓“操纵基因”(operator gene)以及后来发现的所谓“启动基因”(promoter gene)。显然，“操纵基因”和“启动基因”的概念是与顺反子学说相悖的：这两个所谓“基因”的突变拷贝是无法用反式构型中的野生拷贝进行互补的，因为这两种所谓“基因”不编码可在细胞内扩散的产物(蛋白质分子或 RNA 分子)。由于重组 DNA 技术和 DNA 序列分析技术的巨大进展，人们发现所谓“操纵基因”短得可怜，只有二十余 bp。

鉴于上述原因，遗传学家特别是分子遗传学家都陆续剥夺了“操纵基因”和“启动基因”的基因资格，将它们称为 operator 和 promoter 或者 operator site 和 promoter region 等。这种变化在国外早在 70 年代已经开始。例如 1976 年 J. D. Watson 写的《Molecular Biology of the Gene》第3版已干净彻底地剥夺了“操纵基因”和“启动基因”的基因资格。即使是经典遗传学家也悄悄地抛弃了“操纵基因”和“启动基因”的说法。例如 1985 年 M. W. Strickberger 写的《Genetics》第 3 版正文中全部抛弃了这种不恰当的称呼，只在其索引中出现一处 operator gene。这大概是为了照顾一部分人的老习惯。

由上可见，“基因”这一术语使用 80 余年来，其字面并未发生变化，而其概念和涵义却在不断地发生变化，可以说是遗传学中涵义变化最多最快的一个词。到了 1977 年，人们

发现了不连续基因，使得基因的概念并不完全等同于顺反子。基因概念的变迁也必然涉及到其它的遗传学术语。

根据目前人们的认识，基因应该是能够表达和产生基因产物（蛋白质或RNA）的DNA序列。根据产物的类别可以分为蛋白质基因和 RNA 基因（rRNA 基因和 tRNA 基因）两大类；根据产物的功能可以分为结构基因（酶和不直接影响其它基因表达的蛋白质）和调节基因（阻遏蛋白或转录激活因子）两大类。假如按照“有功能的 DNA 片断”或“实现一定遗传效应的核苷酸序列”来定义基因的话，则染色体上处处是“基因”，而且这些“基因”还常常相互重叠，则“重叠基因”和“基因内基因”会多得不可胜数。例如在许多基因内部已发现了阻遏蛋白结合位点（operator），以便形成阴遏状态的噜噗结构（loop）。在经典的乳糖操纵元的 *lacZ* 基因内就有这样的位点。再说，要在染色体上确切地区分“有功能”的序列和“无功能”的序列是很难办到的，这种做法本身或许就是不科学的。这样，即便是一个核苷酸序列已知的染色体，人们还是无法知道这个染色体上到底有多少基因。

关于基因涵义的变迁在国内也早已开始，许多教师和研究人员在他们的授课讲义和论文中已经抛弃“操纵基因”和“启动基因”等不恰当的称呼。不过，在有些遗传学词汇、词典和百科全书中仍然执着于这些不恰当的术语。

2. **Operator——操作子** DNA 分子上阻遏蛋白的结合位点，用以阻止相邻的启动子上转录的起始。B. Lewin 在《Genes》中这样定义操作子："Operator is the site on DNA at which a repressor protein binds to prevent transcription from initiating at the adjacent promoter."
3. **Promoter——启动子** DNA 分子上结合 RNA 聚合酶并形成转录起始复合物的区域，在许多情况下还包括促进这一过程的调节蛋白质结合位点。B.Lewin 在《Genes》中这样定义启动子："Promoter is a region of DNA involved in binding of RNA polymerase to initiate transcription."
4. **Operon——操纵元**（又称操纵子） 是原核生物基因表达和调控的一个完整单元，其中包括结构基因，调节基因，操作子和启动子。B.Lewin 在《Genes》中这样定义操纵元：Operon is a complete unit of bacterial gene expression and regulation, including structural genes, regulator gene(s), and control elements in DNA recognized by regulator gene product(s)."
5. **Regulon——调控元** 一个调节基因的产物同时作用于两个或两个以上的操纵元。这样的调节基因与其控制的操纵元合起来称为调控元。大肠杆菌中最大的调控元为 SOS 调控元。
6. **Transcripton——转录元** 转录单元，即转录作用所涉及的范围。转录元比相应的操纵元要小些。
7. **Replicon——复制元** 复制单元，即从一个复制原点所起始的复制作用所进行的范围。
8. **Transposon——转座元** 转座单元，即能够进行复制并将一个拷贝插入新位点的 DNA 序列单元。
9. **Retroposon——逆座元** 以 RNA 为中间体的转座单元。
10. **Intron——内元**（《科学》杂志称之为内单元） 在原初转录物中通过 RNA 拼接反应而被去除的 RNA 序列或基因中与这种 RNA 序列相应的 DNA 序列。有些基因的内元可以

编码蛋白质(RNA 成熟酶或转座酶),因此称为"内含子"是不恰当的。

12. **Exon——外元**(《科学》杂志称之为外单元) 原初转录物通过 RNA 拼接反应而保留于成熟 RNA 中的序列或基因中与成熟 RNA 相对应的 DNA 序列。几乎全部的蛋白质基因的首尾两个外元都只有部分序列编码氨基酸,也有完全不编码的情况;而在 rRNA 和 tRNA 中的外元当然是不编码的。因此称为"外显子"是不恰当的。

13. **Cistron——顺反元**或**顺反子** 通过顺反测验所确定的遗传单元,大体上相当于一个编码蛋白质的基因。B. Lewin 在《Genes》中这样定义顺反元:"Cistron is the genetic unit defined by the cis/trans test;equivalent to gene in comprising a unit of DNA representing a protein."

14. **Muton——突变子** 发生突变的最小单位,实际上是一个核苷酸对。

15. **Recon——重组子** 发生重组的最小单位,实际上是一个核苷酸对。

16. **Codon——密码子** mRNA 或 DNA 上代表氨基酸或蛋白质合成终止信号的三联体核苷酸。

17. **Enhancer——增强子** 远距离调节启动子以增加转录速率的 DNA 序列,其增强作用与序列的方向无关,与它在基因的上下游位置无关,并具有强烈的细胞类型依赖性。

18. **Silencer——遏制子** 远距离调节启动子以降低转录速率的 DNA。其作用机制与增强子相似,只是效应相反。

19. **Terminator——终止子** 促进转录终止的 DNA 序列,在 RNA 水平上通过转录出的终止子序列形成柄-噜噗结构而起作用。又可分为依赖于 $\rho$ 的终止子和不依赖于 $\rho$ 的终止子两类。

20. **Attenuator——衰减子** 一个受到翻译控制的转录终止子结构。由于翻译作用的影响,衰减子后面的基因或者继续被转录,或者在衰减子处实现转录的终止即产生衰减作用。

21. **Spacer——间隔区** 两种核酸序列之间的任意序列。通常指两个转录单元之间的非转录序列,或者两个基因之间能够转录但不翻译的序列,或者两个 tRNA 基因或两个 rRNA 基因之间在转录后被删除的序列。

22. **Leader——前导序列** 通常指 mRNA 中起始密码子 AUG 之前的序列或与此序列相应的 DNA 序列。

23. **Trailer——拖尾序列** 通常指 mRNA 中终止码密子之后的序列或与此序列相应的 DNA 序列。

24. **Sense strand——有义链** 在双链 DNA 中与转录模板链互补的一条 DNA 链,它与原初转录产物的差异仅在于 DNA 中的 T 变为 RNA 中的 U。有义链又称为编码链(coding strand)。要注意,有义链在过去曾一度被叫做"无义链"(nonsense strand)。

25. **Antisense strand——反义链** 在双链 DNA 中用作转录模板的那条 DNA 链,因此又称为模板链(template strand)。反义链与有义链互补。要注意,反义链在过去曾一度被叫做"有义链"或"编码链"。

26. **Loop——噜噗** 一条核苷酸链盘绕成不规则环形的二级结构。这种二级结构可以由噜噗两端的碱基配对而产生,配对部分称为柄或茎(stem),二者合称为柄-噜噗结构(stem-loop structure),又称为发夹结构(hairpin structure);噜噗也可以由于蛋白质的结合而产生,这样的噜噗既可以是单链,也可以是双链。例如噜噗滚环复制(looped rolling-

circle replication)过程中的噜噗是单链 DNA，而真核生物中调控蛋白质与近启动子复合物作用过程中所形成的噜噗均为双链 DNA。要注意，国内大多把 loop 不恰当地译为"环"，那么噜噗滚环复制就只好叫做使人难于理解的"环形滚环复制"了。

27. **Nick——切刻** 双链 DNA 中一条链上的磷酸二酯键的断裂，而并不缺少碱基或核苷酸。要注意，国内大多把 nick 不恰当地译为"切口"(如叫做单链切口则无含义上的差错，但罗嗦了一点)，还有错误地翻译为"缺口"的。

28. **Nick translation——切刻平移** 在离体试验中，由于 DNA 聚合酶 I 的聚合活性和5′→3′外切酶活性的同时作用，使双链 DNA 上的切刻沿着 5′→3′ 方向由一个位置平行移动到另一个位置。这一作用和过程称为切刻平移。"缺口翻译"是极其错误的译法。这里的 translation 是物理学名词，意为平移或平动的意思。

29. **Isoschizomer——同位酶** 识别位点相同的一组限制性内切酶。要注意，一组同位酶所产生的粘性末端不一定相同。

30. **Isocandamers——同尾酶** 产生相同的粘性末端的一组限制性内切酶。要注意，一组同尾酶的识别位点不一定相同。

31. **Repressor——阻遏蛋白** 与操作子结合的调控蛋白质。对于可诱导操纵元来说，阻遏蛋白本身就是与操作子结合的活性形式；而对于可阻遏的操纵元来说，阻遏蛋白需要与辅阻遏物(corepressor)结合后才能与操作子结合。

32. **Inducer——诱导物** 能够诱导基因表达的小分子物质。对于可诱导的负控制操纵元来说，小分子的诱导物与阻遏蛋白的结合使阻遏蛋白从操作子上解离下来而引起基因的转录；而对于可诱导的正控制操纵元来说，诱导物与无辅基诱导蛋白(apoinducer)的结合才生成有活性的诱导物去促进基因转录。

33. **Selenocysteine——硒半胱氨酸** 是从大肠杆菌直到人类都有的稀有氨基酸之一，是翻译时堂堂正正进入肽链的一个氨基酸，而不是翻译后修饰所产生的。硒半胱氨酸的密码子是 UGA。

34. **Paracodon—— 副密码子** tRNA 分子上决定其携带氨基酸种类的区域。

35. **Transformant——转化体**(也有译为转化子的) 让天然的或人工重组的质粒，噬菌体或病毒进入细胞并能在细胞内复制而繁殖。含有这样的载体的细胞就称为转化体。

36. **Progenote——原始细胞** 生命进化过程中最早出现的细胞叫做原始细胞。由原始细胞进化成古细菌、真细菌和真核生物这三个分支。

37. **Archaebacteria——古细菌** C. Woese 于 1990 年建议改为 archaeotida。

38. **Eubacteria——真细菌** C. Woese 于 1990 年建议改为 bacteriotida。

39. **Procaryotes—— 原核生物** 没有核膜结构的生物，包括古细菌和真细菌。

40. **Eucaryotes——真核生物** 具有核膜结构的生物。

41. **Urcaryotes ——原始真核生物** 由原始细胞进化到单细胞真核生物过程中的中间阶段的生命形态。原始真核生物细胞内没有线粒体，这是它与真核生物的最大区别。

42. **Virion——病毒粒子** 是病毒的细胞外存在形式。根据外膜的有无，可以分为裸病毒粒子(naked virion)和包被病毒粒子(enveloped virion)两大类。

43. **Envelope——外膜** 包被病毒粒子的最外面的脂双层结构，其中含有病毒的大钉蛋白(spike)。在包被病毒粒子的形成过程中，病毒的大钉蛋白首先嵌合于细胞膜中，然后以某

种机制吸引核粒(见以下两条),再才以出芽方式分泌出包被病毒粒子。

44. **Nucleocapsid——核粒** 由病毒的基因组及其衣壳蛋白所组成的结构单元。裸病毒粒子实际上就是核粒。也有人将 nucleocapsid 口语化地译为"壳包核酸"。

45. **Capsid——衣壳** 由病毒自身编码的衣壳蛋白所组成,一般由一种或少数几种衣壳蛋白一再重复而构成。由于这种限制,病毒衣壳只有中空棒状和20面体这样两类形式。

46. **Positive strand RNA virus——正链RNA病毒** 基因组 RNA 能够作 mRNA 或与 mRNA 相同,这样的单链 RNA 病毒称为正链 RNA 病毒。

47. **Negative strand RNA virus——负链 RNA 病毒** 基因组 RNA 进入细胞后必须通过 RNA 复制,产生出互补链作为 mRNA,,这样的单链 RNA 病毒称为负链 RNA 病毒。

48. **Pregenome——前基因组** 在乙型肝炎病毒基因组 DNA 进入细胞后,以负链为模板转录产生全长的正链 RNA,并与病毒编码的逆转录酶和引物蛋白质一起包装到非成熟衣壳中。在这样的衣壳中以正链 RNA 为模板合成全长的负链 DNA,然后正链 RNA 降解,并以其降解片断为引物,以负链 DNA 为模板合成不完整的具有不定长度的正链 DNA。最后添加包装信号,形成完整的衣壳,在分泌过程中获得其外膜,最终成为病毒粒子。上述过程中所产生的全长正链 RNA 则称为乙肝病毒的前基因组。

49. ***cdc* genes——*cdc*基因** 维持细胞分裂周期正常运转的一组基因。在酵母细胞中已发现50种以上的 *cdc* 基因。

50. **Ribozyme——R酶** 全称应该是"核糖核酸质酶",表示这类酶的化学本质是核糖核酸,是 RNA 所具有的催化性质。它仅在化学本质上不同于传统的"蛋白质酶"(proteineous enzymes),而在催化功能及其特点上和蛋白质催化剂却没有什么不同,只不过是在灵活性和多样性方面远不如蛋白质酶。近年来,国内关于 ribozyme 译名的讨论和建议已经很多了。尽管每个人都费尽了脑筋,但都还不能尽如人意。这是近百年来中国文字在新名词翻译方面所面临的困境的继续,因为世界上各种语言之间广泛采用的字母对译方式(transliteration)是中国目前文字所难以采用的。既然如此,就只能按照ribozyme的意义和本质来定译名了。在这方面权威性的论述就是 ribozyme 的发现者 Cech 1986 年的论文"RNA as an Enzyme"(*Sci. Amer*.255:64~75)。Thomas Cech 发现四膜虫 26S rRNA 的自我拼接,以及 L-19 分子对于其它分子的催化性质;Sidney Altman 发现了 RNAaseP 的催化成分是其中的 RNA 即 M1RNA 而不是蛋白质。RNA 催化性质的发现可以说是80年代分子生物学中最重大的发现,因此上述二人获得1989年诺贝尔化学奖。

# 索　引

（说明：索引按条目的前三个字的汉语拼音音序排列；若以英文字母或希腊字母开头，则优先排列）

## A


Ac-Ds 系统　449
*ada* 基因　149
A-DNA　13
ADP 核糖基转移酶　150,302～303
Alu 序列　41,273,462
amber 抑制 tRNA　250
ANT-C　337
AP 内切核酸酶　136,148
AP 位点　148～149
ApppA　326
*att* 位点　321,424～428
α-互补　472～473
α-螺旋　379,382,386～395
α-微管蛋白　370
阿拉伯糖操纵元　293～296
氨甲喋呤与 DHFR 基因扩增　338
暗修复　139


## B


B1 序列家族　41
Bam 岛　360
B-DNA　13
Berk-Sharp 制图法　58
Breathnach-Chambon 规则　233
Britten-Davidson 模型　348～350
BX-C　337
β-半乳糖苷酶　283～286
β-内酰胺酶　272,483
β-微管蛋白　370～371
八核苷酸序列　358,361
靶点倍生　431
半保留复制　84
伴翻译处理　271
半甲基化　347
半乳糖操纵元　291～293
胞嘧啶脱氨氧化　136
保守 A　231
报警素　326
编码链　184
变性　10
表达载体　475
表现型　159
表型抑制　162
别乳糖　286
柄-噜噗结构　201～202
丙糖磷酸异构酶　73
病毒大家族　444
病毒核粒　451
病毒粒子　451
不均等交换　43～44,74
不连续基因　56
不稳定 RNA　210
"不转录间隔区"　359


## C


C 值矛盾　32～33
CI 蛋白　38～39,306～307,380～385
CII 蛋白　306～309,319～320
CIII 蛋白　306
C5-甲基胞嘧啶　17, 151～156, 177～178, 346～348
CAP-cAMP　188,191～192,290～291
*cca* 基因　215
CCAAT 框　194,350
cDNA 克隆　477～478
Chi 结构　410～411
Chi 序列　422～423
$CI^{857}$　475
$C^{ind}$　294～296
*c-onc* 基因　457
*copia* 成分　446～447
*cos* 位点　112
*cos* 质粒　476
Cot 曲线　11～12

$C^{rep}$ 294～296
Cro 蛋白 39,306～307,380～335,389
CTF 355～356
操纵元 34～35,187,283～285
操纵子(见操纵元)
操纵基因(见操作子)
操作子 34～35,187,285
拆分酶 438
插入序列 430～431
缠绕数 22
长末端重复 431
长散布重复序列 461
超螺旋 18,21～28
超螺旋密度 22
超阻遏 283
沉寂匣子 340
持家基因 335～336
重复序列 41
重组热点 422～423
重组修复 142
穿梭载体 473,475
粗面内质网 273
错配矫正酶 138
错配修复 137
错义突变 160
错义抑制 164,167～169

## D

D-嘌噗 86,102～103
*dam* 基因 138,176
DHU 臂 247～248
*din* 基因 146
δ 成分 445～446
DNA
A-DNA 13
B-DNA 13
变性 10
不均一性 7
Chargaff 当量规律 8
*cos* 位点 112
Cot 曲线 11～12
缠绕数 22
超螺旋 18,21～28
超螺旋密度 22
大沟 3,13～15
大沟抵括 19～21
带子模型 22
单链环 DNA 27～28
动基体 7,54
二核苷酸对 8
二级结构 3,13-21
Ⅱ形 DNA 27～28
反式构象 15,17
反向重复序列 7
复性 10～12
复制 84～131
负超螺旋 18,21～28
富含 AT 序列 7,9
构象 13～21
呼吸作用 9
环连 DNA 27～28
碱基对 3,13～14
碱基堆集力 4～5,8
碱基倾角 13～21
紧缠 21,23
可弯曲性 52
两类遗传信息 1
链环数 22
链环数比差 22
磷酸基 6
螺距 3,13～14
螺旋浆扭转效应 19,21
螺旋扭角 13～21
螺旋轴 13～16
扭曲数 22
氢键 3～4,14～15
RNA-DNA 杂交分子 18
Ⅲ形 DNA 27～28
顺式构象 15,17
松缠 21,23
松弛形式 23,25,27
$T_m$ 值 8～10
坍缩 DNA 27
突变 159～182
拓扑缠绕数 22
拓扑异构现象 21～28
拓扑异构酶 23-27
Vinograd 方程 22
V形 DNA 18
卫星 DNA 7,41～45

线形单链DNA 27~28
小沟 3,13~15
小沟抵牾 19~21
溴化乙锭 25~26
修复 136~157
序列分析 2
序列复杂性 11~12
旋转酶 24~25,94~95
一级结构 1
I形DNA 27~28
I°形DNA 27~28
Z-DNA 14~21
Z-DNA结合蛋白 17
杂交 11
主链 3,13~14
DNAase I 192~193
DNAaseI优先敏感性 342~343
DNAaseI超敏感性 343~344
DNA重排
*cin* 基因 443
*gin* 基因 442~443
*H*1和*H*2基因 326
HO内切酶 424~425
*hin* 基因 326
995bp序列及其倒位 326
酵母接合型互变 339~341,424~425
Mu噬菌体G片断及其倒位 442~443
免疫球蛋白基因的生成 341
P1噬菌体C片断的倒位 443
*rh*1基因 326
沙门氏菌鞭毛相变 326,443
DNA复制 84~131
半保留复制 84
半不连续性 95~97
D-噜噗 86,102~103
*dut*⁻ 96~97
dUTPase 96~97,136
大T抗原 112
蛋白质引物 117~119
端粒复制 126
端粒酶 126
复制复杂性 84
复制原点 84~87,100~108
复制体 102
复制元 102,120
复制元族 120
复制终止 116~119
复制调控 128~131
*gam* 基因 112
γ复合体 90
冈崎片断 96~97,116
共价延伸 109~111
滚环复制 86~88,111,422~423
核苷单磷酸激酶 89
核苷二磷酸激酶 89
核小体与复制 126~128
后随链 97,115~116
蝴蝶形分子 124
尖毛虫 126
λ复制起始 105~106
连环分子 88,117
末端蛋白 117
末端冗余 117
NF-2 119
尿嘧啶N-糖基酶 96~97
平锅柄结构 119
起始复合物 105
前体片断的连接 98~99
RFI 113
RFI* 105
RFII 113
Rep蛋白 94,116
SV40 DNA复制 122~126
T7 DNA复制 106~108,117~118
脱氧核苷三磷酸前体 89
*ung*⁻ 96~97
先导链 97,115~116
腺病毒DNA复制 117~120
腺病毒DNA聚合酶 119
腺病毒单链DNA结合蛋白 119
线粒体复制 102~103
胸腺嘧啶核苷酸合成酶 89
引发酶 97~100,113
引发体 105,113
引物 97~100
预引发体 105,113
游仆虫 126
转录激活 98,105
质粒拷贝数控制 130~131
*dna* 基因 100~101

DNA 甲基化
C5-甲基胞嘧啶(CG 二核苷酸对)与基因转录活性 346~348
半甲基化 346~347
各物种甲基化差异 348
甲基化不足 346
去甲基化 347~348
全甲基化 346~347
5-氮胞苷与人工去甲基化 347
C5-甲基胞嘧啶与限制-修饰作用 151~156
C5-甲基胞嘧啶与突变热点 177~178
C5-甲基胞嘧啶与 Z-DNA 17
*dam* 基因 104,131~138
N6-甲基腺嘌呤(GATC 中)与错配修复 137~138
N6-甲基腺嘌呤(GATC 中)与 *oriC* 功能 104
N6-甲基腺嘌呤与限制-修饰作用 151~156
烷基化试剂产生的甲基化
嘌呤 N-7,N-3,O-6 甲基化 149
磷酸基甲基化 149
硫酸二甲酯测定受保护的 DNA 嘌呤碱基 193,287
DNA 聚合酶(原核) 88~93
$\gamma$ 复合体 90
聚合活性 90
链置换 92
内切酶活性 92
Pol Ⅲ* 90,116
切刻平移 92
3′→5′ 外切酶活性 91
5′→3′ 外切酶活性 91
DNA 聚合酶(真核) 120~121,124
DNA 连接酶 93
DNA 螺旋酶 94
DNA 修复 136~150
*ada* 基因 149
AP 内切核酸酶 136,148
AP 位点 148~149
暗修复 139
胞嘧啶脱氨氧化 136
长补丁修复 142
重组修复 142
错配修复 137
错配矫正酶 138
*dam* 基因 138
*din* 基因 146
DNA 糖基酶 148
电离辐射 149
短补丁修复 142
二聚体后起始 143
非标准碱基 147~148
复制修复 136~137
复制后修复 144
干皮性色素沉着 142
光复活酶 139
光修复 139
活存曲线 139
碱基插入酶 148~149
交联 149
LexA 145~146
*mutH,L,S* 138
$N^6$-甲基腺嘌呤 137,151
尿嘧啶 N-糖基酶 136
$O^6$-甲基鸟嘌呤甲基转移酶 149
*phr* 基因 139
PolI 136
切除修复 141
去嘌呤作用 148
RecA 142,145
RecA 蛋白酶活性 145~146
SOS 反应 146
SOS 框 147
SOS 修复 144
蛇形细胞 139
适应反应 149
损伤修复 138~149
*umuC* 144
UV 复活 144
UvrABC 142
烷基化 149
5-甲基胞嘧啶 151
修复内切酶 141
胸腺嘧啶二聚体 138~147
液体保持恢复 139
自身负向控制 147
DNA 旋转酶 94~95
DNA 引发酶(原核) 97~98
DNA 引发酶(真核) 121,124
大 T 抗原 112
大亚基 258

鸟苷酸 208
单点突变 160
颠换 160
电离辐射 149
端粒 55,126
短散布重复序列 461
动基体 7,54
多聚(A)尾巴 208
多聚 ADP 核糖基化 277
多聚核苷酸磷酸化酶 321
多聚核糖体 257
多拷贝抑制 434
多数 tRNA 250

## E

EF-Ts 265,267～268
EF-Tu 265-268
eIF 265
2-氨基嘌呤 172～173
二倍旋转对称 379-380
二核苷酸对 8,346～348
二级结构 3,13～21
二聚环 410
二聚体后起始 143
2′-磷酸 219
2μm 质粒 411
Ⅱ 形 DNA 27～28
Ⅱ 型 tRNA 基因 215
二元复合物 197

## F

φ29 119
φ×174 36～37,113～114
翻译 247～277
  Alu 序列 273
  β-促脂解激素 277
  β-内啡肽 277
  伴翻译处理 271
  粗面内质网 273
  促黑激素 277
  促肾上腺皮质激素 277
  蛋白质合成起始 262
  蛋白质合成速度 262
  多聚 ADP 核糖基化 277
  多顺反子 mRNA 的翻译 269
  EF-Ts 265,267
  EF-Tu 265～268
  eIF 265
  二脂酰甘油 276
  发夹结构(蛋白质) 271
  翻译后处理 271,276～277
  反面高尔基体 274
  泛素 270
  泛素化 277
  泛素羧端水解酶 270
  非蛋白质锚 275～276
  分泌蛋白质 273
  分泌贮备小泡 274
  分拣 274
  GTP 与翻译 263～268
  核糖体 257～262
  核糖体受体蛋白 273～274
  肌醇三磷酸酯 276
  甲基化 277
  磷酸化 277
  磷脂酶 C 276
  N-甲酰-甲硫氨酰-tRNA 262
  内质网 274
  7SL RNA 273
  起始密码子 262
  起始复合物 264
  起始 tRNA 262
  起始因子 263、264
  溶酶体蛋白质 273
  Shine-Dalgarno 序列 264
  释放因子 267
  双功能 mRNA 269
  顺面高尔基体 274
  Ts 循环 267
  tRNA 247～257
  $tRNA_f^{Met}$ 262
  $tRNA_M^{Met}$ 262
  $tRNA_i^{Met}$ 262
  糖基化 274,277
  糖基-磷脂酰肌醇锚 276
  坞蛋白 273～274
  细菌内膜 271
  细菌外膜 271
  信号补丁 274
  信号识别颗粒 273

信号肽 271
信号肽酶 271,274
延伸 265
延伸因子 265~266
乙酰化 277
越膜 271
真核生物蛋白质合成起始 265
真核生物起始密码子的"上下文" 269
中间高尔基体 274
终止与肽链释放 267~269
周质 271
主流机制 274~275
转氨基酸作用 277
转肽与 EF-Tu 265~266
转位与转位因子 EF-G 267
转运小泡 274
翻译水平的调控 316~326
ApppppA 326
*att* 位点 321
报警素 326
CⅡ 与溶原生长 306~309
蛋白质合成自体调控 321~324
单链 RNA 噬菌体 317~318
多基因 mRNA 中各基因产物比例 321
多聚核苷酸磷酸化酶 321
f2 噬菌体 317~318
返回调控 320
反义 RNA 316~317
附着蛋白 317~318
核糖体蛋白质合成 322
核糖体蛋白质 L1 的自体调控 324
*int* 基因 319~321
空转反应 325
λ 噬菌体 319~321
micF 316
micQ 316~317
micRNA 316
mRNA 二级结构 317~318
mRNA 寿命 319~321
MS-2 噬菌体 317~318
魔斑Ⅰ和Ⅱ 325~326
内切酶与 mRNA 寿命 321
OmpC 和 OmpF 316
PaQ 316~317
ppGpp 和 pppGpp 325~326
Qβ 噬菌体 317~318
起始 AUG 316
R17 噬菌体 317~318
RNA 复制酶 317~318
RNAaseⅡ 321
RNAaseⅢ 320
*rel* 基因 325
*rnc*⁻ 突变 320
S-D 序列 316,324
*sib* 位点 320~321
3′外切核酸酶 319
释放因子 RF2 的合成 321~322
松弛突变体 325
严谨反应 325~326
严谨因子 325~326
移框窗口 322
移框作用 321~322
衣壳蛋白 317~318
终止子强度与 mRNA 稳定性 319
翻译后处理 271,276~277
翻译扩增 369
反馈抑制 297
反密码子 249
反密码子臂 247~248
反式拼接 236~237
反式作用 284
反式作用因子 354~356
反向遗传学 178
反义链 184
反义 RNA 316~317,435
返回调控 320
返座 444
返座元 444
泛素 270,277
非标准碱基 147~148
非病毒返座元 460~464
非蛋白质锚 275~276
非条件突变 161
非洲爪蟾 109,338,369
非自主成分 449
分泌蛋白质 273
分泌型载体 475,483
分泌贮备小泡 274
分拣 274
分子进化 71~75,78,225~226

封闭性启动子复合物 197
丰富 mRNA 333-335
辅阻遏物 283
辅助病毒 456
负超螺旋 18,21～28
负控制 281
负控制因子 386
负链 RNA 病毒 185
-10 序列 188～189
-35 序列 189～190
附加臂 247～248
附着蛋白 317～318
附着点 424
副密码子 250～251
复合转座元 431～436
复性 10～11,39～40
复制 84～131
复制重组 405
复制后修复 144
复制修复 136～137
复制元融合 438

## G

G 片断 442～443
G4 DNA 复制 113
*gag* 基因 451
*gal* 操纵元 291～293
*gam* 基因 423
GC 框 354～355
GCD2 基因 362
GCN2 基因 362
GCN4 基因 362
GTP 263～268
干扰素 483～486
感受态细胞 473
感受位点 348～350
冈崎片断 96～97
高度重复序列 41～42
共合体 437～441
共价延伸 109～111
共进化 75
共转换 413
古细菌 76～77
固醇类激素受体 357
光修复 139～141
滚环复制 86～88,111,422～423

## H

*hfl* 基因 307
*hin* A 基因 307
*hin* 基因 443
HMG 蛋白质 342～343
hnRNA(P) 363
HO 内切酶 424
Holliday 中间体 408～411,423
核苷单磷酸激酶 89
核苷二磷酸激酶 89
核糖核酸质酶 225
核糖体 257～262
　A 位点 261
　春日霉素 259
　大亚基 258
　多聚核糖体 257
　EF-Tu 位点 262
　28S rRNA 258
　核糖体数目 257
　核糖体装配 259
　核糖体突变 260
　活性位点 260～263
　mRNA 结合位点 261
　P 位点 261
　*sad* 突变体 259
　16S rRNA 258
　肽基转移酶活性位点 261
　5S rRNA 258
　5S rRNA 位点 261
　小亚基 258
　转位因子 EF-G 结合位点 262
核糖体蛋白质合成 322～324
核小体 48～54,126～128,341
核心酶 90,186
核心序列 O 427
后随链 97,115～116
呼吸作用 9
琥珀型 160
还原病毒 238,450～460
还原转座元 444
回复突变 160～161
回收载体 474～475
迴文序列 7,152

活存曲线 139
活动散布遗传成分 238
活跃匣子 310

## I

$i^{-}$突变体 289~290
$i^{-d}$突变体 290
IF1 263~264
IF2 263~264
IF3 263~264
*int* 基因 309,319~321,427
IPTG 473
IS 430
IS 共合体 438
IS 组件 432
$i^{s}$ 突变体 289~290
$i^{q}$ 突变体 289~290

## J

肌醇三磷酸酯 276
基因 488
  *ada* 149
  *ara BAD* 293~296
  *ara C* 293~296
  *c*I,*c*II,*c*III 38,204,306~309
  *cca* 215
  *cdc* 369
  *cin* 443
  *dam* 138,176
  *din* 146
  *dna* 100~101
  DNA 聚合酶基因 90
  *dut* 96~97
  *env* 451
  $\phi\times174$ 基因 36
  *gag* 451
  *galETK* 291
  *gam* 112,423
  GCD2,GCN2,GCN4 362
  *gin* 442~443
  H1,H2 326
  *hinA* 307
  *hin* 326,443
  *his3* 181
  *hfl* 307
  *int* 309,319~321,427
  *lacI* 284~290
  *lacZYA* 284
  *lexA* 145~146
  *mut* 138,176
  *N* 203,204
  *O* 106
  *P* 106
  *phr* 139
  *pol* 451
  *pyrBI* 314
  *Q* 204
  *rad52* 424
  *recA* 146
  *recBCD* 421
  *red* 423
  *rel* 325
  *rh1* 326
  *rho* 202
  *rnc* 320
  *rpoA,B,C,D* 186,202,206
  *rpoH* 302
  *rrn* 36,62~63,211
  *tnpR* 440~441
  tRNA 基因 64
  *trpEDCBA* 296
  *umuC* 146
  *ung* 96~97
  *ura3* 181
  *uvrABC* 146
  *xis* 309
  串联重复基因 61~63
  II 型 tRNA 基因 215
  鸡卵清蛋白基因 57
  5S RNA 基因 63
  细胞器基因 64
  I 型 tRNA 基因 215
  珠蛋白基因 59~61
  组蛋白基因 61~62
基因表达调控的分子机制 377~401
  $\alpha$ 螺旋 379,382,386~395
  $\alpha$螺旋—转角—$\alpha$螺旋 387~393
  $C_2/C_2$ 型锌指 386,395
  $C_2/H_2$ 型锌指 386,394
  Cro 蛋白的三维结构 389

- Cro 蛋白与 $O_{R3}$ 的作用 383
- DNA 结合结构域 386～387,396
- DNA 序列的对称性 379～380
- 蛋白质相互作用的复杂性 398
- 二倍轴对称 379
- 二倍旋转对称 379～380
- 富含谷酰胺的转录激活区 396
- 富含脯氨酸的转录激活区 396
- 负调控因子 386
- 共享磷酸 382
- 果蝇 Antp 同形异位结构域氨基酸序列 393
- 滑动假说 398
- 基因表达调控的定义 377
- 基因表达调控中的三类作用 377
- 接力假说 398
- 抗红霉素基因 400
- Leu 拉链 386,395
- Leu 拉链与二聚体形成 395
- *lac* 操纵元的双启动子结构 399～400
- λ Cro 蛋白 380～385,389
- λ 早期操纵元的操作子结构 381
- λ 阻遏蛋白 380～385
- λ 阻遏蛋白的结构域 381～382
- λ 阻遏蛋白结合的协同效应 382,397
- λ 阻遏蛋白的负调控作用 383
- λ 阻遏蛋白的正调控作用 382～384
- λ 阻遏蛋白 N 端结构域的结构 387
- λ 阻遏蛋白二聚体与操作子结合模型 388
- λ 原噬菌体的诱导和裂解生长 380,385
- 两类遗传信息及其差异 377～379
- 啮噗假说 396～397
- 螺丝对称 379
- 拧转假说 397
- 平移对称 379
- 起阻遏蛋白作用的 RNA 聚合酶 399
- RNA 聚合酶 Ⅱ 396
- RNA 开关系统 400～401
- RecA 蛋白酶活性 380,385
- 色氨酸衰减子 400
- 识别螺旋 390～393
- 识别螺旋交换试验 390～393
- 噬菌体 P22 390～393
- 噬菌体 434 390～393
- 酸性激活区 382～383,396
- TATA 结合因子 TFD 396
- TFⅢA 394～395
- 调控蛋白质对称性 379～380
- 调控密码 379
- 同形异位蛋白质 393
- U1 snRNA 401
- 锌指结构 386,394～395
- 真核生物的 α 螺旋—转角—α 螺旋 393
- 正调控因子 386
- 正负调控的相对性 398～400
- 转录激活区 396
- 转录因子 Oct1 和 Oct2 393
- 转录因子 Pit 393
- 阻遏蛋白同时与两个操作子结合 396～397

基因定位 67～72
- 接合定位 68～69
- 染色体步行 70～71
- 染色体跳跃 71～72
- 遗传交换定位 67～68

基因丢失 337

基因工程 469～486
- α-互补 472～473
- β-内酰胺酶基因 271,483
- 表达载体 475
- $CI^{857}$ 温度敏感突变体 475
- cDNA 克隆 477～478
- *cos* 质粒 476
- 穿梭载体 473,475
- 促进基因工程的重大发现 469
- 放射标记的 DNA 探针 478
- 非融合型载体 475
- 分泌型载体 475,483
- 复制原点 471,473
- 干扰素基因工程 483～486
- 感受态细胞 473
- 回收载体 474～475
- IPTG 473
- 基因工程制造微量蛋白质步骤 479～480
- 基因组文库 476
- 抗菌素抗性基因 472
- 克隆载体 472～473
- *lacI^q* 473
- *lacZ* ΔM15 473
- *lac* 启动子 475
- λ 噬菌体 476
- M13 系列 473

墨迹法 478
粘性末端 152,470
鸟枪法 478
$R_L$ 和 $P_R$ 启动子 475
pBR322 471~472,476
pEX 系列 475
pKK 系列 475
pUC 系列 472~473
平齐末端 152
人生长激素基因工程 483~486
融合型载体 475
T-DNA 480~481
Ti 质粒 480~481
*tac* 启动子 475
*trp* 启动子 475
同尾酶 152
同位酶 152
土壤杆菌 480
X-gal 473
限制性内切酶 152~153,469~470
信号肽 271,475
胰岛素基因工程 482~485
乙型肝炎基因工程疫苗 486
杂交 180,476,478~479
载体 471~476
质粒扩增 472
植物基因工程 480~481
自主复制序列 ARS 473
基因家族 59~60
基因扩增 109~111,338
基因型 159
基因转换 45,411~413
基因簇 60~61
基因组 32
极化元 413
加性饱和杂交试验 335~336
甲基化不足 346
甲基化 DNA 17,104,137~138,149,151~156,177~178,193,287,346~347
甲基化 RNA 207~208,213,258~259,400~401
碱基插人 160
碱基插入酶 148~149
碱基对 3,13~14
碱基类似物 171~173
碱基缺失 160
碱基替代 160
交互重组 408~409
交换固定 45
交联 149
脚印法 192
接受位点 348~350
解旋酶活性 198,421
介入序列 54
紧缠 21,23
进化钟 75
居民 DNA 151
均一化作用 45

## K

Klenow 片断 192
开放性启动子复合物 188,197
抗红霉素基因 400
抗药性基因 431~432
抗终止蛋白 203~206,306
抗终止机制 205~206
可弯曲性 52
可诱导操纵元 282
可诱导顺式成分 356
可阻遏操纵元 282
克隆载体 472~473
空转反应 325
控制成分 443
枯草杆菌 298~301

## L

L-19 223~226
*lac* 操纵元 34,187~188,283~291,399~400
*lacI* 284
$lacI^q$ 473
*lacO* 284,287~290
$lacO^C$ 287,290
*lacUV-5* 192,475
*lacZ M15* 473
λ 噬菌体
*attB* 和 BOB′ 序列 424~428
*attL* 和 BOP′ 序列 426~427
*attP* 和 POP′ 序列 424~428
*attP* 的蛋白质结合位点 428
*attP* 的核心序列与 Tn3 *res* 位点的比较 441
*attR* 和 POB′ 序列 426~427

α螺旋一转角一α螺旋 382,387～390
CI 蛋白 38～39,306～307,380～385
CⅡ 蛋白与溶原生长 306～309,319～320
CⅢ 蛋白 306
Chi 突变* 422～423
Cro 蛋白 39,306～307,380～385,389
Cro 蛋白的三维结构 389
Cro 蛋白与 $O_{R3}$ 的作用 389
*cy* 突变体 309
重组与噬菌体成熟 422～423
返回调控 319～320
复制起始 105～106
复制原点 106
复制终止 117
附着点 424
*gam* 基因与滚环复制 111,422～423
滚环复制 111,422～423
共享磷酸 382
*hfl* 基因 307
*himA* 基因 307
核心序列 O 427
*int* 基因 309,319～321,427
基因 *O* 和 *P* 产物与复制起始 105～106
抗终止蛋白 203～206
抗终止机制 205～206
裂解生长与溶原生长的调控 306～309
烈性突变 309
MOI 值 307
micQ 316～317
免疫性物质 39
N 蛋白 203～206,306
*nut* 206
*nutL* 205
*nutR* 205
逆转刺激因子 FIS 426
*ori* 序列 106
$P_E$ 307
$P_I$ 307,309,320,427
$P_L$ 189,475
$P_M$ 307
$P_R$ 189,475
$P_{aQ}$ 316～317
Q 蛋白 204～205
*qut* 205
切除酶 426～428
清亮噬菌斑 309
*red* 基因 423
*rnc⁻* 320
溶原化 38～39,307～309,424,427
*sib* 位点 320
识别螺旋 390
酸性激活区 382～383
*θ* 复制 97,111,422～423
UV 复活 144～145
晚期操纵元 203
*xis* 基因 309,427
线形分子和环形分子的转变 37～38
限制和修饰模式 151
遗传图 38
营养耗竭 307
右向早期操纵元 203
原噬菌体 424
原噬菌体的诱导和裂解生长 39,147,321,380,385
早期操纵元的操作子序列 381
整合的分子机制 427～428
整合寄主因子 IHF 426～428
整合酶 309,319～321,426～428
整合与切除 424～427
转录激活 105
阻遏蛋白操纵元 203
阻遏蛋白的结构域 381～382
阻遏蛋白 N 端结构域的结构 387
阻遏蛋白结合的协同效应 382,397
阻遏蛋白二聚体与操作子结合模型 388
阻遏蛋白的负调控作用 383
阻遏蛋白的正调控作用 382～384
左向早期操纵元 203
作为克隆载体 71～72,476
Leu 拉链 386,395
Lex A 145～146
LINE1 461
类表皮生长因子结构域 72
类 IS 430,432
离体定向诱变 178
利福霉素 198
连环分子 88,117
联会 405
联会丝复合体 405～406
链环数 22

烈性突变　309
磷酸化　277,345
磷酸基　6
磷酸基甲基化　149
磷脂酶 C　276
硫酸二甲酯　175,193
噜噗结构　293,396～397
螺旋桨扭转效应　19,21
螺旋扭角　13～21
螺旋轴　13～16

## M

M13　112,180,473
M1 RNA　225
Meselson-Radding 模型　406～408
mic F　316
mic Q　316～317
mic RNA　316
MOI 值　307
mRNA 二级结构　317～318
mRNA 翻译起始的调控　369～370
mRNA 前体的选择性拼接　364～368
mRNA 寿命　319～321
mRNP 复合物　363
Mu 噬菌体　176,441～443
*mut* 基因　138,176
芒脚突变　336
帽子　207～208
密码子　249～257
免疫性物质　39
魔斑　325～326
末端蛋白　117
末端倒转重复　430,432,434,439
末端冗余　117
模板链　184

## N

N 蛋白　203～206,306
$N^6$-甲基腺嘌呤　104,137～138,151～156,208
$N^7$-甲基鸟苷酸　207～208
NNG　173,175
nu 结构　420
内部引导序列　221
内共生　77～79
内含子(见内元)
内元　32～33,36,216～237,364～368,461
逆转录　238～240
逆转录酶　238～240,445,451,464
粘性末端　152,470
鸟枪法　478
尿嘧啶 N-糖基酶　96～97,136
拧转假说　397
扭曲数　22

## O

OmpC 和 OmpF　316
*onc* 基因　451,458
*ori* 序列　106
*oriC*　102～105
*oxi3* 基因　231

## P

P 成分　447～449
P 位点　261
P1 噬菌体　443
P16 和 P23　212～213
$P_{aQ}$　316～317
$P_B$　307
*phr* 基因　139
$P_I$　306,309,320,427
$P_{IN}$　434
$P_{OUT}$　434
$P_L$ 和 $P_R$　189,475
$P_M$　307
PolI　136
Poly(A)尾巴　209～210
Poly(A)聚合酶　209～210
PolyC 聚合酶活性　225
Pribnow 框　188～189
平齐末端　152
平移对称　379

## Q

Q 蛋白　204～205
Qβ 噬菌体　317～318
*qut*　205
启动子　34,187～197,350～351
启动子上升突变　161,192
启动子下降突变　161,192
歧化　74

起始核苷酸位点 198
起始密码子 262
起始复合物 105,264
起始 tRNA 262
起始因子 263～264
嵌合剂 175～176
羟胺 174～175
切除修复 141
亲核进攻 225
轻度重复序列 41
氢键 3～4,14～15
清亮噬菌斑 309
去分枝酶 232
去嘌呤作用 148

## R

R 酶 225
*rad52* 基因 424
RecA 146,380,385,404～405,410,414～423
RecBCD 420～423
RecF 途径 423
*red* 基因 423
*rel* 基因 325
Rep 蛋白 94,116
*res* 位点 440～441
RF I 113
RFI* 105
RF II 113
*rh1* 基因 326
*rho* 基因 202
RNAase II 321
RNAase III 213,216,320
RNAase D 214～216
RNAase H 239,452～453
RNAase P 216,225
RNA 饱和杂交试验 33
RNA 成熟酶 56,228～231
RNA 重排酶 223
RNA-DNA 杂交分子 18
RNA 复制 227～228
RNA 复制酶 317～318
RNA 甲基化
  N7-甲基鸟嘌呤与帽子 0 207～208
  N6-甲基腺嘌呤与帽子 1 207～208
  核糖 2′-0-甲基化与帽子 1,2 207～208
  核糖 2′-0-甲基化真核 45S rRNA 前体转录后处理 213,258～259
  原核 16S,23S rRNA 甲基化 258
  原核 23s rRNA 红霉素结合位点上 A 的甲基化 400～401
RNA 聚合酶(原核) 184～187
RNA 聚合酶(真核) 187,396
RNA 聚合酶 I 启动子 359～360
RNA 聚合酶 II 启动子 193～195,350
RNA 聚合酶 III 启动子 195～197,360～361
RNA 开关系统 400～401
RNA 末端腺嘌呤核苷酸转移酶 206,208
RNA 拼接 216～237
  Breathnach-Chambon 规则 233
  保守 A 231
  补骨脂素交联 235
  $C_5$ 蛋白质 225
  *cob* 基因 228
  *cob*⁻突变点分布 228～230
  长基因 228
  第 I 类内元 217,221～231
  第 II 类内元 217,231～233
  地中海贫血病 233
  短基因 228
  2′-磷酸 219
  反密码子与 tRNA 拼接 217
  反式拼接 236～237
  核基因 mRNA 217,233～236
  核基因拼接过程 235～236
  核糖核酸质酶 225
  环形内元 223
  酵母细胞色素 b 基因 221
  酵母细胞色素 C 氧化酶亚基 a 基因 221
  酵母线粒体 217
  L-19 223～226
  $M_1$RNA 225
  内元 217
  内部引导序列 221
  *oxi3* 基因 231
  PolyC 聚合酶活性 225
  拼接体 234
  亲核进攻 225
  去分枝酶 232
  R 酶 225
  RNAaseP 225

RNA 成熟酶 228～231
RNA 重排酶 223
RNA 异构酶 223
RNA 催化与生命进化 225～228
RNA 复制 227～228
scRNP 233
Sm 蛋白 235
snRNP 233
SP6 启动子 232
3′ 拼接点 217
四膜虫 217,221～227
四膜虫大 rRNA 223
四膜虫大核 223
四膜虫微核 223
T4 内元 228
tRNA 基因 217
tRNA 拼接 217～221
tRNA 半分子 219
套索结构 232,236
U1snRNP 234～236
U2snRNP 235
U5snRNP 235
U4/U6snRNP 235
外元 217
5′ 拼接点 217
小外元 236～237
胸腺嘧啶核苷酸合成酶 228
Y 形拼接产物 237
中部核心结构 217,221
转酯反应 223,232～233
锥虫 236
RNA 寿命 210
*rnc* 基因 320
Rot 曲线和 Rot $^1/_2$ 333～335
*rpoA,B,C,D* 186,202,206
*rpoH* 基因 302
*rrn* 基因 36,62～63,211
rRNA 和 tRNA 的转录后处理 210～216
不稳定 RNA 210
*cca* 基因 215
Ⅱ型 tRNA 基因 215
甲基化核苷酸 213
间隔子 214
P16 212～213
P23 212～213
RNA 寿命 210
RNAaseD 214,216
RNAase P 214,216
RNAase P4 216
RNAaseⅢ 213,216
*rrn* 操纵元 211
45S 前体 213
47S 前体 214
tRNA 核苷酸转移酶 215
稳定 RNA 210
Ⅰ型 tRNA 基因 215
染色体 31,33～34,39～55
染色体步行法 70～71
染色体交叉 406
染色体跳跃法 71～72
染色体外分子 63
染色质 47～48,341～348
热震惊基因 301～302
融合型载体 475
溶酶体蛋白质 273
乳糖操纵元 34,187～188,283～291,399～400

## S

S-腺苷甲硫氨酸 154～155
S-腺苷高半胱氨酸 155
*sad* 突变体 259
scRNP 233
S-D 序列 264,310,324
Sextama 框 189～190
*sib* 位点 320
SINEs 461
snRNP 210,233
SOS 反应 146
SOS 框 147
SOS 修复 144
SP1 354～355
ssb 蛋白 101,112～115,419
SV40 120～126,350～353
σ32 302
σ70 302
σ 操纵元 302
三元复合物 197
色氨酸操纵元 296～297
沙门氏菌鞭毛相变 443
上游启动子成分 350

少数 tRNA 250
奢侈基因 335～336
识别螺旋 390～393
释放因子 267
释放因子 RF2 合成的调控 321～322
适应反应 149
衰减子 310～315
双功能 mRNA 269
顺式作用 284
顺式作用成分 284,350～354
四膜虫 217,221～227
松缠 21,23
松驰突变体 325
松驰形式 23,25,27
酸性激活区 382～383,396
损伤修复 138～149

## T

T4 噬菌体 228,302～303
T7 噬菌体 106～108,117～118,303～306
TATA 框 194,350
TATA 结合因子 TFD 396
TATA 样序列 361
T-DNA 480～481
TFⅢA 394～395
Ti 质粒 480～481
$T_m$ 值 8～10
Tn 共合体 438
TnA 转座元家族 430,438～441
tRNA 247～257
amber 抑制 tRNA 250
氨酰基 tRNA 合成酶 250
*cca* 基因 215
DHU 臂 247～248
第Ⅰ类 tRNA 248
第Ⅱ类 tRNA 248
动力学校对 251
多数 tRNA 250
Ⅱ型 tRNA 基因 215
反密码子 249
反密码子臂 247～248
副密码子 250～251
附加臂 247～248
构象校对 251
化学校对 251
还原病毒引物 tRNA 451～452
接受臂 247～248
碱基堆集 249
L 形结构 248
密码子 249
密码子简并性 249
密码子使用频率 251～254
密码通用性 254～255
密码的进化 256～257
三叶草结构 247～248
少数 tRNA 250
TψC 臂 247～248
tRNA 半分子 219
tRNA 分子内氢键 249
tRNA 丰富度 251～254
tRNA 核苷酸转移酶 215
tRNA 拼接 217～221
同功 tRNA 250～251
线粒体密码 254～255
Ⅰ型 tRNA 基因 215
抑制 tRNA 165～169
摇摆配对 249～251
原始副密码子 251
*trp* 操纵元 296～297
*trp* 衰减子 310～314
Ty 成分 445～446
坍缩 DNA 27
糖基化 274,277
糖基-磷脂酰肌醇锚 276
套索结构 232,236
体细胞中分离现象 64
天门冬氨酸转氨甲酰酶 314
调节基因 285
调控 281
调控元 35
同尾酶 152
同位酶 152
同形异位现象 336～337
同形异位框 337,393
同源重组 404～424
8 字形中间体 409～410
Chi 结构 410～411
Chi 序列 422～423
Chi 序列与重组热点 422～423
Chi 序列与细菌接合 423

重组热点 422～423
重组小瘤 406
单链侵入 406
定义 404
二聚环及其转变 410
2μm 质粒 411
非同源链交换与 SOS 诱变 419
分支迁移 407～408
*gam* 基因 423
共转换 413
HO 内切酶 424
Holliday 中间体及其拆分 408～411,423
基因转换 411～413
基因转换的极性 413
极化元 413
减数分裂 405
减数分裂后分离 412
交互重组 408～409
酵母 MAT 序列转换 424
解旋酶活性与再旋酶活性 421～422
λ 噬菌体的成熟与重组 422～423
联会 405
联会丝复合体 405～406
链交换的方向 418～419
链同化 407
链置换 406
嘧啶切除 407
Meselson-Radding 链转移模型 406～408
nu 结构 420
切断 406
RecA 404～405,410,414～423
RecA 的性质 415～416
RecA 促进链交换 418～419
RecA 促进 Holliday 中间体形成 420～421
RecBCD 420～423
RecF 途径 423
*red52* 基因 424
*red* 基因 423
染色体交叉 406
ssb 蛋白 419
同源联会的步骤 416～417
兔耳结构 422
外切核酸酶 V 421
细菌接合 414
细菌转化 413
相外同源重组 45,405
异构化 407
异源双链 408～409,411～413
子囊菌 411～412
子囊孢子 411～412
阻遏蛋白 Sir 424
突变 159～182
表型抑制 162
错义突变 160
错义抑制 164,167～169
*dam* 基因 176
DNA 分析遗传学 178
单点突变 160
多点突变 160
定点突变 179
颠换 160
定义与分类 159～161
2-氨基嘌呤 172～173
反向遗传学 178
放射性探针 180
非条件突变 161
寡聚核苷酸 179
*his3* 基因 181
琥珀型 160
化学修饰 173～175
回复突变 161
回复突变频率 162
基因间抑制突变 162,164～171
基因间间接抑制突变 170～171
基因内抑制突变 162～164
基因转移 180
间接抑制突变 162
碱基插入 160
碱基类似物 171～173
碱基缺失 160
碱基替代 160
离体定向诱变 178
硫酸二甲酯 175
M13 噬菌体 180
Mu 噬菌体 176
*mut* 基因 176
NNG 173,175
启动子上升突变 161,192
启动子下降突变 161,192
嵌合剂 175～176

羟胺 174～175
乳石型 160
渗漏突变 161
T4 DNA 聚合酶 180
条件型突变 161
同义突变 160
同源重组 181
酮式和烯醇式互变异构 171
突变频率 160
突变热点 177～178
*ura3* 基因 181
烷基化试剂 175
温度敏感突变 161
温度敏感抑制突变 169
无声突变 161
无义突变 160
无义抑制 164～167
5-甲基胞嘧啶 178
5-溴尿嘧啶 171～172
吖啶橙 175
吖黄素 175
亚硝酸 174
诱发突变 160
移框突变 160
移框抑制 164,169
抑制突变 161
抑制基因 164
抑制增强突变 169
原黄素 175
原位回复突变 161
杂交 180
增变基因 176
赭石型 160
正向突变 161
直接抑制突变 162
致死突变 160
中性突变 161
转换 160
紫外线 176
自发突变 160
组成型突变 161
组成型表达 161
拓扑异构现象 21～23

## U

U1snRNP 231～236,401
U2snRNP 235
U5snRNP 235
U4/U6snRNP 235
U6 基因 360
*uvrC* 基因 146
*ung* 基因 96～97
*ura3* 基因 181
UV 复活 144

## V

V 形 DNA 18
Vinograd 方程 22
*v-onc* 基因 457～459

## W

W 复活 144
外膜 451
外显子(见外元)
外元 56,71～75
烷基化试剂 175
网织红细胞 369
尾丝蛋白 442～443
位点特异性重组 424～428
*attB* 和 BOB′ 序列 424～428
*attL* 和 BOP′ 序列 426～427
*attP* 和 POP′ 序列 424～428
*attP* 上的蛋白质结合位点 428
*attR* 和 POB′ 序列 426～427
附着点 424
核心序列 O 427
酵母 MAT 序列的转换 424～425
λ 噬菌体 DNA 的整合与切除 424～427
λ 噬菌体整合的分子机制 427～428
逆转刺激因子 FIS 426
切除酶 Xis 426～428
原噬菌体 424
整合寄主因子 IHF 426～428
整合酶 426～428
整合和切除时 *int* 基因的不同表达方式 427
卫星 DNA 7,42～47
温度敏感突变体 161
稳定 RNA 210

无辅基诱导蛋白 282
无辅基阻遏蛋白 283
无声突变 161
无义突变 160
无义突变极性效应 202
无义抑制 164~167
5-氮胞苷 347
5-甲基胞嘧啶 17,151~156,177~178
5-溴尿嘧啶 171~172
坞蛋白 273~274

## X

*xis* 基因 309,427
硒半胱氨酸 1,256
细胞器 64~67
先导链 97,115~116
显性 284
腺病毒 117~120
线粒体 65~67,78~79
线粒体密码 254~255
限制酶 150~157
限制与修饰 150~157
相外同源重组 45,405
小沟 3,13~15
小沟抵牾 19~21
小外元 56,236~237
小亚基 258
协同效应 191
新霉素磷酸基转移酶 435
锌指结构 386,394~395
信号补丁 274
信号识别颗粒 273
信号肽 271,275
信号肽酶 271,274
胸腺嘧啶二聚体 138~147
胸腺嘧啶核苷酸合成酶 228
修复 136~157
修复内切酶 141
修饰酶 150~157
溴化乙锭 25~26
旋转酶 24~25,94~95

## Y

Y 形拼接产物 237
延长核苷酸位点 198
延者 99
延伸控制颗粒 206
延伸因子 265~266
严谨反应 325~326
摇摆配对 249~251
野生型 159
液体保持恢复 139
I 形 DNA 27~28
I⁰ 形 DNA 27~28
I 型 tRNA 基因 215
一致进化 75
遗传重组 404~464
　同源重组 404~424
　位点特异性重组 404~405,424~428
　遗传重组的定义 404
　遗传重组的分类 404
　异常重组 404~405
　转座 404~405,429~464
遗传漂变 75
异构化 407
异肽键 345
异源双链 408~409,411~413
抑制基因 164
抑制 tRNA 165~169
引发酶 97~100,113
引发体 105,113
隐蔽 mRNA 369
隐蔽卫星 DNA 42
隐性 284
有义链 184
诱导 282
预引发体 105,113
原病毒 450~456
原核生物 31
原黄素 175
原始副密码子 251
原始细胞 76~78
原噬菌体 39,147,321,380,385,424
原位回复突变 161
越膜 271

## Z

Z-DNA 14~21
杂交 11,180,476,478~479
载体 70,471~476

赭石型突变 160
真核基因表达 332~341
ANT-C 337
氨甲喋呤引起的 DHFR 基因扩增 338
BX-C 337
沉寂匣子 340
持家基因 335~336
非洲爪蟾 rDNA 扩增 338
丰富 mRNA,优势 mRNA 333~335
丰富度 335
回放试验 333
活跃匣子 340
基因丢失 337
基因扩增 111,338
加性饱和杂交试验 335~336
酵母接合型互变 339~341
芒脚突变 336
免疫球蛋白基因生成 341
区域特异性转录因子 337
RNA 饱和杂交试验 333
RNA 驱动的杂交 333
$Rot^1/_2$ 333~335
双小染色体 338
奢侈基因 335~336
同形异位框 337
同形异位突变 336
同形异位现象 336
稀少 mRNA,复杂 mRNA 333~335
真核基因表达与染色质结构 341~348
半甲基化 347
CG 二核苷酸对 346~348
DNAase I 优先敏感性 342~343
DNAase I 超敏感位点 343~344
DNA 超螺旋与 DNAase I 超敏感点 344
各种真核 DNA 甲基化的差别 348
核小体 341
HMG 蛋白质 342~343
甲基化不足 346
甲基化和去甲基化 346~348
rRNA 基因的包装比 341
SV40 微染色体 342,344
5-氮胞苷 347
异肽键 345
转录起始与核小体 342
组蛋白的修饰 345~346
组蛋白的泛素化 345~346
组蛋白的甲基化 345
组蛋白的磷酸化和去磷酸化 345
组蛋白的乙酰化 345
真核基因表达的转录调控 348~363
Britten-Davidson 模型 348~350
Bam 岛 360
八核苷酸序列 358,361
"不转录的间隔区" 359
CCAAT 框 194,350
CTF 355~356
促进 rRNA 基因有效转录的因素 360
反式作用因子 354~356
反式因子与顺式成分的作用 356
GC 框 354~355
GCD2 基因 362
GCN2 基因 362
GCN4 基因 362
感受位点 348~350
固醇类激素受体 357
酵母通用氨基酸控制 361~362
结构基因 348~349
接受位点 348~350
可诱导顺式成分 356
噜噗结构 354
免疫球蛋白基因表达 358~359
能产型复合物 353
7SK 基因 360
启动子 350~351
RNA 聚合酶 I 启动子 359~360
RNA 聚合酶 III 启动子的多形性 360~361
rRNA 基因转录特点 359~360
SP1 354~355
SV40 增强子 350~353
三种 RNA 聚合酶启动子的共通性 361
上游启动子成分 350
顺式作用成分 350~354
TATA 框 194,350
TATA 样序列 361
U6 基因 360
增强子 350~353,361
增强子成分 351~354
增强子元 351~354
增强子与启动子关系 354
正调控因子激活 357

转录的翻译控制 361~362
综合者基因 348~350
阻遏蛋白失活 357
组织特异性顺式成分 358
真核基因表达的转录后调控 207~210, 216~237,363~368
不同 3′末端型 367~368
不同 5′末端型 367
二甲基精氨酸 363
反式拼接 236~237
hnRNA 的选择性加工运输 363
hnRNP 颗粒 363
互斥型 365
肌钙蛋白 T 基因 365~366
鸡肌红蛋白 367
降钙素相关基因 367~368
mRNA 前体选择性拼接 364~368
mRNP 复合物 363
内部拼接点型 365
内元滞留型 367
RNA 拼接 216~237
3′多聚 A 尾巴 209~210,363
5′帽子 207~208,363
匣子型 365
纤维结合蛋白 365
选择性拼接 365
组成型拼接 364
真核基因表达的翻译调控 368~371
α-微管蛋白 370
β-微管蛋白 370~371
*cdc* 基因 369
蛋白质合成的自体调控 370~371
翻译扩增 369
家蚕丝心蛋白的合成 369
mRNA 翻译起始的调控 369~370
mRNA 稳定性 368~369
S6 磷酸化 369
四氨基酸信号 370
生长因子 369
网织红细胞 369
隐蔽 mRNA 369
影响 mRNA 稳定性的因素 368~369
招募因子 369
终端分化细胞 369
真核生物 31,76~77
真细菌 76~77
正、负调控的相对性 398~400
正控制 281
中度重复序列 41
终端分化细胞 369
终止子强度与 mRNA 稳定性 319
周质 271
主链 3,13~14
主流机制 274~275
转录 184
编码链 184
柄-噜噗结构 201~202
*E.coli* RNA 聚合酶 186
二元复合物 197
反义链 184
封闭性启动子复合物 197
负链 RNA 病毒 185
核心酶 186
迴文序列 200~202
解旋酶活性 198
近基因 202
开放性启动子复合物 188,197
抗终止蛋白 203~206
λ 阻遏蛋白操纵元 203
利福霉素 198
利福平 198
模板链 184
N 蛋白 203~206
NusA 206
NusB 206
*nut* 位点 206
*nutL* 205
*nutR* 205
Poly(A)序列 206
Q 蛋白 204~205
*qut* 205
起始核苷酸位点 198
全酶 186
RNA 合成步骤 185
RNA 合成速度 199
RNA 合成特征 185
RNA 聚合酶 184~187
RNA 末端腺嘌呤核苷酸转移酶 206,208
ρ 因子 200~206
*rho* 基因 202

*suA* 突变 202
三氯醋酸沉淀法 185
三元复合物 197
晚期操纵元 203
无义突变极性效应 202
有义链 184
右向早期操纵元 203
延宕 199
延长核苷酸位点 198
延伸控制颗粒 206
远基因 202
再旋酶活性 198
真核生物 RNA 聚合酶 187
真核生物转录终止 206
正链 RNA 病毒 184
终止子 200～206
转录起始 197
转录起始点的选择 198
转录终止 200
左向早期操纵元 203
转录产物的修饰 207～210
戴帽酶 208
多聚(A)尾巴 208
mRNP 210
帽子 207～208
帽子 0 207～208
帽子 1 207～208
帽子 2 207～208
帽子功能 208
帽子结合蛋白 208
$N^6$-甲基腺嘌呤 208
$N^7$-甲基鸟苷酸 207～208
Poly(A)聚合酶 209～210
RNA 末端腺苷酸转移酶 208～210
snRNP 210
U1snRNA 210
转录的操纵元调控 281～297
AraC 的正、负控制功能 294～296
*araI* 294～296
$araO^C$ 294～296
*araO2* 294～296
阿拉伯糖操纵元 293～296
β-半乳糖苷酶 283～286
半乳糖表异构酶 291
半乳糖操纵元 291～293
半乳糖双启动子 292～293
半乳糖激酶 291
半乳糖转移酶 291
本底组成型合成 286
别乳糖 286
不可诱导 283
$C^{ind}$ 294～296
$C^{rep}$ 294～296
cAMP-CAP 290～291
操纵元 283～285
操作子 285
超阻遏 283
反馈抑制 297
反式作用 284
辅阻遏物 283
负控制 281
*galP1*, *P2* 292
$galO_E$, $O_I$ 292～293
互补试验 284
IPTG 283～284
$i^-$突变体 289～290
$i^{-d}$ 突变体 290
$i^s$ 突变体 289～290
$i^t$ 突变体 289～290
结构基因 285
可诱导操纵元 282
可阻遏操纵元 282
*lac* 阻遏蛋白 286～290
*lacI* 284
*lacO* 284,287～290
$lacO^C$ 突变 287,290
噜噗结构 293
乳糖操纵元 283～291
色氨酸操纵元 296～297
顺式作用 284
*trpEDCBA* 296
*trpO* 296～297
*trpR* 296～297
调节基因 285
无辅基诱导蛋白 282
无辅基阻遏蛋白 283
显性 284
隐性 284
诱导 282
诱导物 282～283

正控制 281
阻遏 283
阻遏蛋白 281
转录的时序调控 298 309
ADP核糖基转移酶 302～303
CⅠ蛋白 306～307
CⅡ蛋白 306
CⅢ蛋白 306
Cro蛋白 306～307
*cy*突变体 309
*hfl*基因 307
*himA*基因 307
*int*基因 309
枯草杆菌中的σ亚基更迭 298～301
枯草杆菌噬菌体SPO1 298～301
λ噬菌体两种生长周期的选择 306～309
烈性突变 309
MOI值 307
$P_E$ 307
$P_I$ 309
$P_M$ 307
清亮噬菌斑 309
*rpoH*(*htpR*) 302
热震惊反应 301～302
热震惊基因 301
热震惊基因启动子 302
溶原状态的建立 307～309
σ32 302
σ70 302
σ操纵元 302
T4 RNA聚合酶的修饰 302～303
T4 Mot蛋白 303
T7 RNA聚合酶 303
T7 RNA聚合酶启动子 305
T7的三组转录产物 303～306
T7的生长周期 304
*xis*基因 309
营养耗竭 307
整合酶 309
转录的翻译调控 310～315
不依赖ρ因子的终止子 310～312,314～315
核苷酸合成酶的衰减子 314～315
*lac*操纵元 310
连续的色氨酸密码子 310～313
*pyrBI*操纵元 314
*pyrBI*前导序列 315
色氨酰-tRNA合成酶突变 310
衰减子的普遍性 313
衰减子的生物学意义 314
*trp*操纵元 310～313
*trp*前导肽 310～312
*trp*前导序列 310～312
*trp*衰减子 310～313
*trp*衰减子二级结构 311～312
天门冬氨酸转氨甲酰酶 314
暂停位点 314～315
转录与翻译的耦联 311,314
转位因子EF-G 265～267
转运小泡 274
转酯反应 223,232～233
转座(细菌) 404～405,429～443
靶点倍生和同向重复 431
拆分酶 438
拆分酶基因*tnpR* 440
插入序列 430～431
长末端重复 431
蛋白质1的转座功能和顺式作用 435～436
蛋白质2的转座抑制和反式作用 435～436
多拷贝抑制 434
反义RNA 435
非对称模型 438
复合转座元 431～436
复合转座元的起源 433
复制重组 405
复制元融合 438
G片断的倒位 442～443
G(+)走向 442～443
G(－)走向 442～443
*gin*基因 442～443
共合体 437～441
共合体的拆分 437～441
共合现象 437～438
*him*基因 443
*hin*,*gin*,*cin*和*tnpR*四种基因产物的比较 443
IS 430
IS组件 432
IS共合体 438
995 bp片断的倒位 443
抗药性基因 431～432

类 IS 430,432
类 IS 组件 432
Mu 噬菌体 441~443
Mu 噬菌体的特点 442
末端倒转重复 430,432,434,439
内部拆分序列 440
$P_{IN}$ 434
$P_{OUT}$ 434
区域性优先 430,439
*res* 位点 440~441
沙门氏菌鞭毛相变 443
Tn 共合体 438
TnA 转座元家族 430,438~441
TnpR 的阻遏蛋白活性 440~441
Tn3 438~441
Tn3 共合体拆分与 λ 噬菌体重组的比较 441
Tn5 433,435~436
Tn9 432
Tn10 432,434~435
Tn903 433
Tn1681 432
尼丝蛋白 442~443
新霉素磷酸基转移酶 435
中心序列和臂 431
转座 404~405,429
转座机制 436~439
转座酶 431,434~435,438
转座酶的顺式作用 434
转座酶基因 *npAt* 440
转座频率和切除频率 430
转座热点 434
转座与进化 429
转座元 429
转座(真核生物) 443~464
Ac-Ds 系统 449
Alu 序列 462
病毒大家族 444
病毒核粒 451
病毒粒子 451
病毒转录激活因子基因 451
*copia* 成分 446~447
c-*onc* 基因 457
长末端重复 $U_3RU_5$ 451
长散布重复序列 461
δ 成分 445~446
短散布重复序列 461
*env* 451~452,457
FB 成分 446~447
返座 444
返座机制 463~464
返座基因 460~461
返座假基因 460~461
返座元 444
非病毒大家族 444
非病毒返座元 460~464
非自主成分 449
分类 444
辅助病毒 456
*gag* 451
还原病毒 450~460
还原病毒的癌基因 456~459
还原病毒基因组结构 451~452
还原病毒 RNA 到原病毒 DNA 的转变 451~455
还原病毒与转座元的关系 459~460
还原转座元 444
控制成分 443
LINEs 461
LINE1 461
逆转录酶 445,451,464
*onc* 基因 451
*onc* 基因分类 458
P 成分 447~449
P 成分与杂种不育 448
*pol* 451
7SL RNA 462
前胰岛素原 I 基因 460
强致癌还原病毒 456
RNA 聚合酶Ⅱ转录产物返座元 461
RNA 聚合酶Ⅲ转录产物返座元 461~463
弱致癌还原病毒 456
SINEs 461
Ty 成分 445~446
Ty 成分的增强子 445
U2 snRNA 假基因 460
U3 序列 451
U5 序列 451
v-*onc* 基因 457
v-*onc* 基因的致癌机制 458
v-*onc* 基因来自 c-*onc* 基因 459